IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 June 2022, accepted 24 June 2022, date of publication 5 July 2022, date of current version 19 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3188649

SURVEY

Machine Learning in Precision Agriculture: A Survey on Trends, Applications and Evaluations Over Two Decades

SARAH CONDRAN[1], MICHAEL BEWONG[1], (Member, IEEE), MD ZAHIDUL ISLAM[2], LANCELOT MAPHOSA[3,4], AND LIHONG ZHENG[1], (Member, IEEE)

[1]School of Computing, Mathematics and Engineering, Charles Sturt University, Wagga Wagga, NSW 2650, Australia
[2]School of Computing, Mathematics and Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia
[3]NSW Department of Primary Industries, Wagga Wagga, NSW 2650, Australia
[4]Australian Department of Agriculture, Water and the Environment, Canberra, ACT 2601, Australia

Corresponding author: Sarah Condran (scondran@csu.edu.au)

This work was supported by Charles Sturt University.

ABSTRACT Precision agriculture represents the new age of conventional agriculture. This is made possible by the advancement of various modern technologies such as the internet of things. The unparalleled potential for data collection and analytics has resulted in an increase in multi-disciplinary research within machine learning and agriculture. However, the application of machine learning techniques to agriculture seems to be out of step with core machine learning research. This gap is further exacerbated by the inherent challenges associated with agricultural data. In this work, we conduct a systematic review of a large body of academic literature published between 2000 and 2022, on the application of machine learning techniques to agriculture. We identify and discuss some of the key data issues such as class imbalance, data sparsity and high dimensionality. Further, we study the impact of these data issues on various machine learning approaches within the context of agriculture. Finally, we identify some of the common pitfalls in the machine learning and agriculture research including the misapplication of machine learning evaluation techniques. To this end, this survey presents a holistic view on the state of affairs in the cross-domain of machine learning and agriculture and proposes some suitable mitigation strategies to address these challenges.

INDEX TERMS Agriculture, digital farming, intelligent agriculture, machine learning, precision agriculture, precision farming.

I. INTRODUCTION

The term precision agriculture (*aka* digital farming or intelligent agriculture) has been used to describe the incorporation of various technologies into traditional farming practices, to improve agricultural productivity and sustainability [1]–[3]. Modern technologies such as the Internet of Things (IoT) paves the foundation of precision agriculture that enables the minimisation of human labour and cost as well as improving agricultural productivity. IoT generates large volumes of data which can be used for practices such as crop monitoring or disease detection. The analysis and interpretation of this data enable the understanding of relationships between various agricultural factors such as soil characteristics and climatic variables. This facilitates timely and informed decision making and planning [1]. Typically, such decision support systems have been used in bio-security applications, quality assurance, farm and resource management, and land usage [4]. Machine learning (ML) plays a central role in these decision support systems by modelling the complex patterns that may exist in the data. Figure 1 illustrates a typical precision agriculture scenario where decision support may be used. The figure represents a three-tiered precision agriculture architecture adapted from [5]. The first tier is the physical layer which represents the hardware that is in proximity with the farm elements. This layer is mostly

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

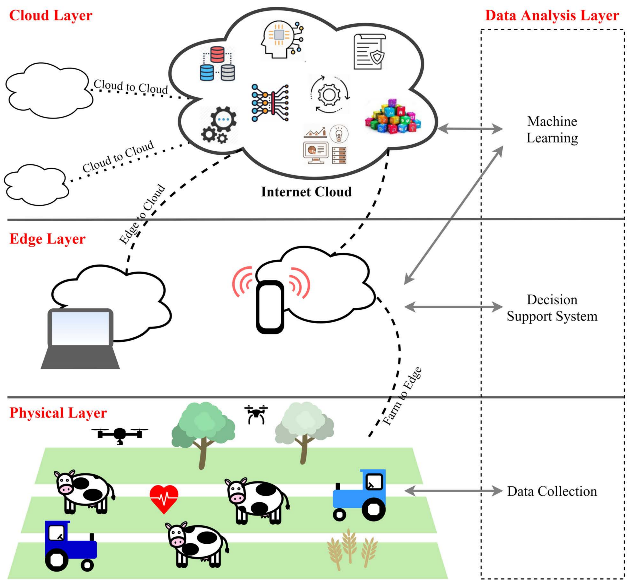

FIGURE 1. The precision agriculture ecosystem.

made of sensors and actuators. The second tier which is the edge layer represents the computing devices which may be used either on premise or remotely to analyse or interpret the data collected in the physical layer. This layer is often made of devices with low to medium computing resources. The third tier is the cloud layer which represents the IT infrastructure that supports the storage, processing and analysis of the data. The cloud layer is often made of computing resources with high throughput and storage capacity. The cloud layer supports the edge layer for decision making about the farm based on data collected by the physical layer.

The use of IoT within the agricultural domain is increasing due to its ability to support increased production capabilities and analytics. The application of IoT can be categorised into monitoring, predict/forecast, control and logistics/documentation tasks [6]–[8] with devices occurring at different levels including field, vehicular, aerial, and satellite [7], [9], [10]. Monitoring provides automatic data collection of various parameters including soil data such as moisture and chemistry, crop data including leaf area and plant height, and weather data including rainfall and humidity [7], [9], [10]. For example, [11] monitored crop storage for moisture and temperature to reduce crop damage, while [12] monitored the crop using a wireless sensor network (WSN) to protect the potato crop against a fungal disease. Data collected through the devices along with historical data is used to predict or forecast, and to provide knowledge and tools to support decision making. Often ML is used to provide this analysis. Work by [13] used soil sensors to monitor soil moisture and processed the data with a variation of a decision tree to facilitate irrigation planning. IoT devices known as actuators provide remote control to modify the process or environment. Common applications within agriculture include irrigation, fertiliser and pesticide application, illumination and access [6]. [14] applied IoT to optimise greenhouse growing conditions such as humidity, luminosity and irrigation. IoT has enabled the flow of information from producers to consumers and include post farm processes such as handling, packaging, transportation, and distribution. In the works of [15], the conditions during transport are assessed to improve value through reduced food spoilage, energy costs and increased product quality.

There have been several propositions on how ML can be applied to the agricultural domain for decision support. In particular, much attention has been given to the crop, water, and soil sectors. However, there is a lack of consensus within the research community on the best practices in employing ML techniques. This is no trivial issue as agricultural data possess some of the most challenging characteristics of datasets encountered within the ML domain. That is, agricultural data are often imbalanced, sparse, and riddled with noise [16]–[23]. For example, in disease prediction, data are often imbalanced leading to poorer classification models, and sometimes overly optimistic model performance estimates [24], [25]. Further, the commentary on critical aspects such as model trustworthiness is very limited. Trustworthiness is essential for the uptake of ML driven precision agricultural practices by the farming community, and its effective application by researchers. However, the ML literature within the agricultural domain suggests little consideration has been made to model trustworthiness. Indeed, a larger emphasis has been placed, sometimes erroneously, on model accuracy.

At the same time, the survey literature to the best of our knowledge, does not provide a holistic perspective on these aforementioned challenges. They mostly focus on only one aspect at a time. Recent surveys that reviewed the use of ML within specific agricultural applications namely yield prediction [26]–[28] and honey origin prediction [29]. While the works of [30], [31] reviewed the use of specific algorithms within agriculture namely support vector machines and Bayesian networks. Other recent surveys focused on reviewing the use of ML within agriculture [1], [32], [33], and soil science [34]. Of these works [26], [28], [32], [33] covered performance metrics while [34] mentioned model interpretability. Consequently, a literature review that juxtaposes these challenges with the advances in ML in a more holistic manner is warranted.

This work has three contributory parts. The first part identifies the trends within the machine learning and agriculture (ML & Agriculture) literature. We find that there has been an increase in the number of publications within the cross-domain of ML & Agriculture over the last two decades, while the type of ML techniques used has changed.

The second part reviews how ML has been applied within the agricultural domain. We identify and contextualise the commonly used ML techniques within agriculture. Techniques of note are decision tree (DT), k-nearest neighbour (kNN), random forest (RF), support vector machines (SVM), and neural networks (NN). This contextualisation focuses on the strengths and weaknesses when the aforementioned data issues are encountered as well as their interpretability within this context. Next, some of the key data issues encountered within agricultural data are covered. Issues include class imbalance (CI), sparsity, and high dimensionality (HD). Lastly, a brief overview of some applications of ML within the agricultural domain is provided.

The third part investigates the evaluation of ML models in the agricultural setting. Methods used to assess ML models are often geared towards the measurement of classification accuracy [35]–[37]. Classification accuracy serves the dual purpose of identifying the optimal classifier for a situation and to assess the ability of the classifier to complete a task. We argue that, within the context of agriculture, other important factors such as trustworthiness through interpretability are equally important, and in some cases preferable. For example, [38] reviewed the barriers to adopting ML within the agricultural domain and identified lack of model trustworthiness as the fundamental barrier for the adoption of ML. Our work reviews the various model evaluation metrics and discusses their suitability for the ML techniques within the context of agriculture.

The rest of this paper is organised as follows. Section 2 presents the methods, defines the research questions and the search process. Section 3 reports the results from the systematic literature review to address the research questions. Section 4 presents some open research challenges within ML & Agriculture research. Finally, Section 5 presents a conclusion.

II. METHOD

This study provides a multifaceted view of the application of ML within the agricultural domain. To facilitate this study, research questions were developed along with a search process.

A. RESEARCH QUESTIONS

The overarching motivation of this study was to review *whether ML is being effectively applied to the domain of agriculture*. As such, this study addresses the following research questions and sub-questions:

- **RQ1** What are the literature trends in the cross-domain of ML & Agric?
  - **RQ1.1** What are the publication trends over time?
  - **RQ1.2** What are the thematic trends of publications over time?
- **RQ2** How is ML being applied within the agricultural domain?
  - **RQ2.1** What are common ML algorithms used?
  - **RQ2.2** What data issues are encountered within agricultural datasets?
  - **RQ2.3** What are some applications of ML within agriculture?
- **RQ3** What methods are used to evaluate ML model performance, and are they appropriate for the agricultural domain?
  - **RQ3.1** What are the main methods to evaluate ML model performance?
  - **RQ3.2** What metrics are used to measure classification accuracy within the agricultural domain?
  - **RQ3.3** What is the importance of using model interpretability to evaluate ML performance?

RQ1 aims to analyse the trends of ML & Agriculture over the last 20 years. Through a two-tiered approach focusing on the volume of works published and the content of the works. RQ2 explores the application of ML within the agricultural domain, identifying the common ML algorithms and data issues that are present within agricultural datasets impacting the selection of ML algorithms. RQ3 addresses how the performance of ML models are evaluated.

B. SEARCH PROCESS

The review of the literature within the cross-domain of ML & Agriculture was varied and interrogated at differing depths. The overall literature search methodology is depicted in Figure 2. To identify the literature trends over the last 20 years, both the publication and thematic trends were used. The publication trends identified the number of papers within the cross-domain published for each year. The data were obtained from Google Scholar with the search term *Machine learning AND Agriculture*. Next, data for identifying thematic trends were obtained through searching Google Scholar and Web of Science with the search term *Machine learning AND Agriculture*. The title and abstract of each of these papers were collected. Further, an in-depth analysis of the literature was carried out to identify the evaluation trends and ML techniques. The data was obtained from Google Scholar with variations of the search terms *agriculture*, *machine learning* and *classification*. An exclusion criteria were created to limit the scope of this work. Some exclusion criteria include the use of regression and image analysis.

III. RESULTS

In this section, the results of the study are discussed. First, the literature trends over the last 20 years are presented (RQ1). Next, the results of how ML is applied within the agricultural domain is discussed (RQ2). Lastly, how ML model performance is evaluated is assessed, providing experimental examples on the various methods (RQ3).

A. LITERATURE TRENDS (RQ1)

1) PUBLICATION TRENDS (RQ1.1)

The publication trends over 20 years were obtained through Google Scholar searches. Using the search term *Machine learning AND Agriculture*, the number of returned search results were recorded. These trends are depicted in Figure 3. While agricultural research remained relatively constant through time, there has been an increase in ML research. The cross-domain of ML & Agriculture saw over a 100-fold increase, indicating an upward trend of multidisciplinary

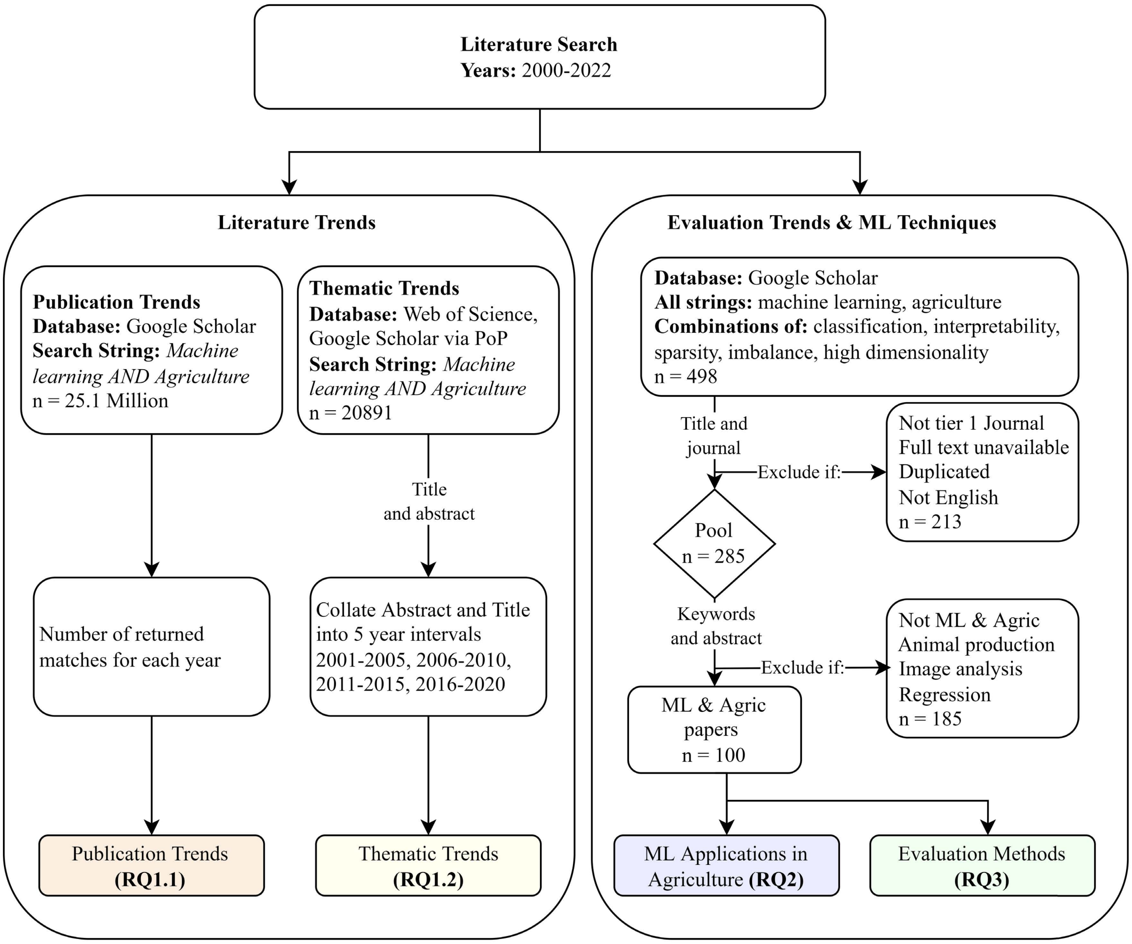

FIGURE 2. Literature review methodology.

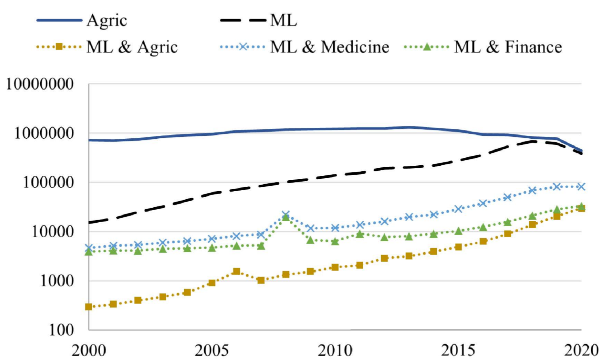

FIGURE 3. Publications over time, obtained through Google Scholar[1].

research. Other domains such as medicine and finance corroborate this trend of multidisciplinary research.

2) THEMATIC TRENDS (RQ1.2)

The literature from the last 20 years covering over 20 thousand papers was analysed to identify the themes and how they changed over time. Literature was collected through Google Scholar [2] and Web of Science. Both search engines used the search phrase *Machine learning AND Agriculture* to identify the literature. The title and abstract of over 20 thousand papers were collated into 5 year intervals i.e. 2001-2005, 2006-2010, 2011-2015, and 2016-2020. We adopted a 5 year interval since we believe it is a sufficient time period for a research area to fully mature. Natural language processing techniques such as stop word removal, lemmatization and stemming were applied to the corpus. Finally, word clouds were created using the top ten frequent words of the current and the previous time interval. Figure 4 comprises of four word clouds, one for each five year interval. The words clouds identified a change in ML techniques over time. "SVM" and "DT" decreased while "RF" and "deep [learning]" increased, this indicates a shift in the types of techniques used. Further, "regression" increased in frequency in the last ten years. This shift in ML techniques can be attributed to a change in the problems addressed and types of data used. Expanding on the problems addressed, there was an increase

[1] Search terms are combinations of Agriculture, Machine Learning (ML), Medicine and Finance.

[2] The corpus of literature was identified by the software *Publish or Perish*(PoP) which collected the first thousand results of each year.

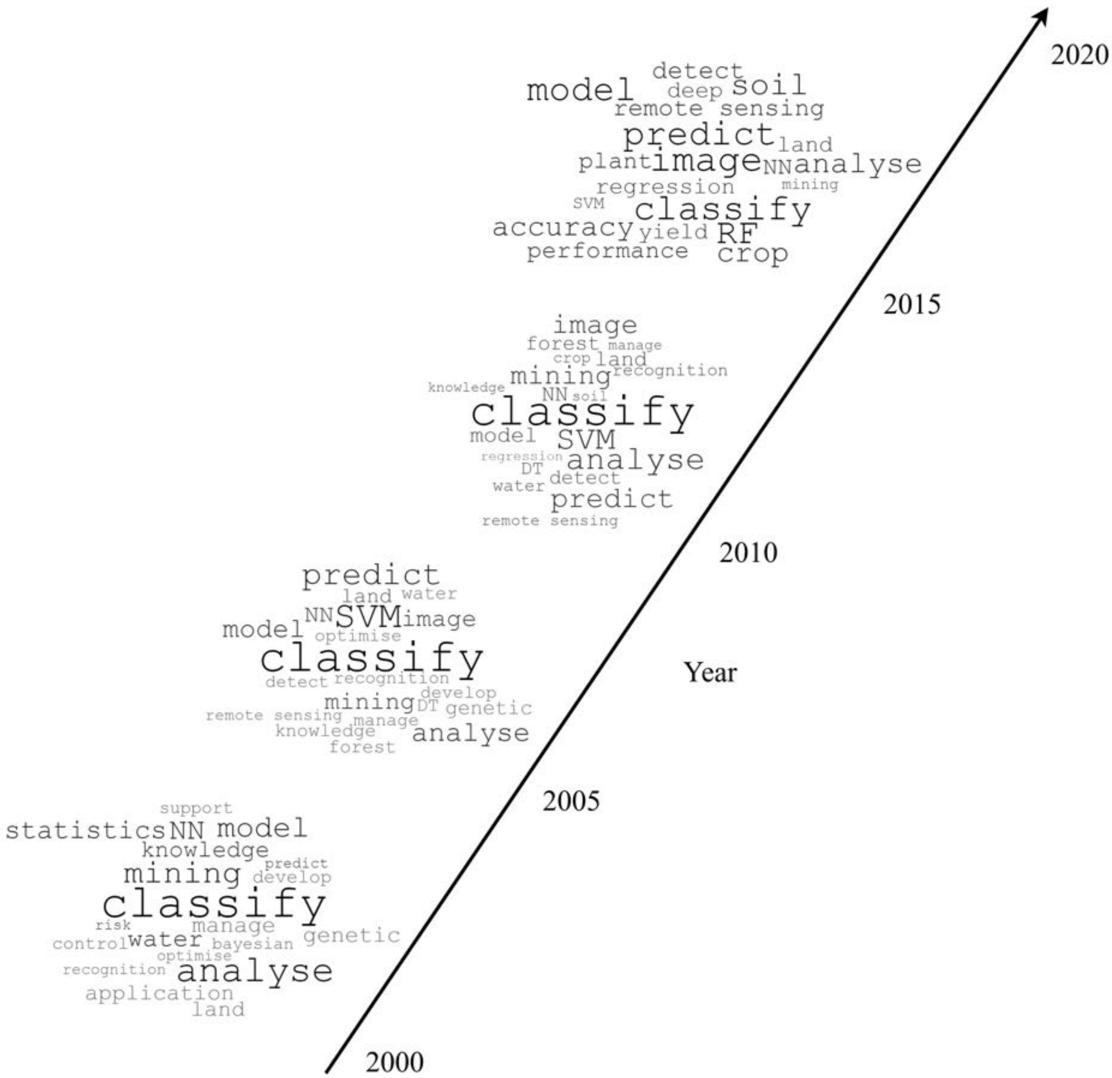

FIGURE 4. Thematic trends of ML & Agriculture literature over two decades.

in frequency of words ''predict'', ''detect'' and decrease in ''mining'', ''statistics'' and ''knowledge''. This indicates the change in the application of ML from an overall gathering of statistics and knowledge to a range of tasks including prediction and detection problems. This can be attributed to the wider range of agricultural problems that ML is used for. The word cloud further identified a shift in agricultural sectors applying ML. Words such as ''yield'' and ''plant'' increased while ''water'' and ''genetic'' decreased. These words indicate the spread of ML throughout the agricultural sectors. Lastly, words such as ''remote-sensing'' and image'' increased in frequency. This indicated the rise in various data collection methods and types, of note remote sensing technology such as IoT.

B. ML IN THE AGRICULTURAL DOMAIN (RQ2)

To understand how ML is used within the agricultural domain, 100 papers were analysed to identify the common ML techniques and the agricultural problems ML is used to solve.

1) ML ALGORITHMS (RQ2.1)

The commonly used ML algorithms within the agricultural domain are decision tree (DT), k-nearest neighbour (kNN), random forests (RF), support vector machine (SVM), and neural networks (NN). These techniques can be categorised by their level of interpretability i.e. the inherent characteristic within the learning technique that allows a user to explain the resulting classifier [39]. The two major categories are interpretable white box models and non-interpretable black box models. It is important to note that this categorisation is not strictly binary. That is, if a white box is complex enough it becomes non-interpretable, while a typically black box model if simple enough can be interpretable [39].

The review of the literature found a change in type of ML used: transitioning from white box to the less interpretable

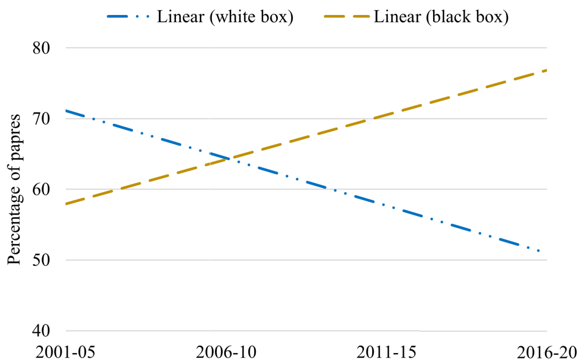

FIGURE 5. Proportion of papers using white and black box ML techniques.

high accuracy black box techniques. Figure 5 shows the proportion of papers using white or black box techniques for each time interval. Of note, RF has significantly increased in occurrence over the last 20 years to become the most prevalent method.

A white box model is characterised by a high level of inherent interpretability of the classifier. Coincidentally, these are also characterised by relatively lower classification accuracies. These techniques, which are usually rule-driven, are common within the agricultural setting for their simplicity. DT and kNN are the most common of these techniques. DT is a tree like structure which aims to generate correct classifications by partitioning the dataset to improve the homogeneity of the classes at each subsequent level. While DT is praised for its simplicity, it can be prone to producing locally optimised solutions [39]. Also it can be overfitted to the data, generating complex models that are not applicable to unseen data, and so decrease interpretability [39]. DT has been applied within the agricultural domain to predict disease in cherry fruit [40] and the risk of disease in boysenberry [41], to forecast climatic events such as draught [42], and the prediction of soil fertility [43]. kNN is a simple but robust technique which relies on the most prevalent class within the closest k points to classify an instance. Similar to DT, kNNs are a non-parametric classification technique. kNN is known to have relatively higher computational costs since it is a lazy learner: that is, no model is built. Rather a prediction is made for each new test instance by comparing it to historical data every time. However, the use of data structures such as KD trees can improve efficiency. With the shift to real time analysis within the agricultural domain, other techniques such as DT could be more efficient. A few applications of kNN within the agricultural domain include the prediction of frost events that will impact crops [44] and the forecast of wheat yields [20], [24].

Algorithms labelled as black box are characterised by inherently low interpretability. This makes them less amenable to interpretation albeit with relatively higher classification accuracies. These techniques are usually based on complex mathematical functions, often with non-linear relationships between the features of the dataset and the target variable [39]. RF, SVM and NN are the commonly used black box models within the agricultural domain. RF can produce highly accurate classifiers which are robust even in the presence of outliers and noisy data. RF are an ensemble of DT, which mitigates overfitting and produces more robust classifiers when compared to individual DTs. However, this model is hard to visualise and interpret, thus many within the ML research community consider it black box [45]. Some applications of RF within the agricultural domain include nutritional diagnosis in peach orchards [17] and the prediction of soil classes [18]. SVMs are effective yet simple techniques which identify a hyper plane to distinguish classes. SVMs are generally difficult to interpret due to the kernel transformations that obfuscates the relationships between the features and the target variable [46]. Applications of SVM within the agricultural domain include the classification of soil quality [47] and type [48] and the classification rice origins [49]. NN simulate the neural network systems of the human brain. NN develop links between input features and the target variable through the iterative adjustments of the node weightings to minimise classification error [50]. The multiple non-linear transformations between the input features and target variable makes it non-interpretable. That is, it is impossible to establish the relationship between input features and the target variable. Applications of NN include the prediction of wheat yields [51] and the tracing of honey origins [16], [52].

2) DATA ISSUES (RQ2.2)

Agricultural data are often riddled with data issues, including CI, sparsity and HD [16]–[23]. The ML techniques previously discussed can behave differently in the presence of these data issues. A dataset is said to have CI when there is a disparity in the representation of one or more classes in the dataset [53]. For instance training data used to predict diseased shellfish farms in [54] had an overwhelming imbalance ratio of 9:1 of healthy to diseased shellfish farms. This has been shown to be quite a common occurrence in agricultural data with 76 % of the papers reviewed presenting some degree of CI [3]. Figure 6 presents a breakdown on the CI by the level of severity: the categories are arbitrarily defined by ratios with imbalance less than 3:1 as slight, between 3:1 and 10:1 as moderate and greater than 10:1 as severe. The impact of CI can be significant and even considered to be the major obstacle to building accurate classifiers [55]. Several applications of ML techniques in the agricultural setting focus on the maximisation of accuracy. However, these techniques often assume equal class distributions. Unfortunately, as seen in Figure 6, this assumption is rarely realistic. A classifier built with a class imbalanced dataset can produce high testing accuracy, but perform poorly in real world applications [53], [56]. This problem is further exacerbated by the small sample sizes

[3]Interestingly of the 100 papers reviewed 34 did not include any description of class distribution.

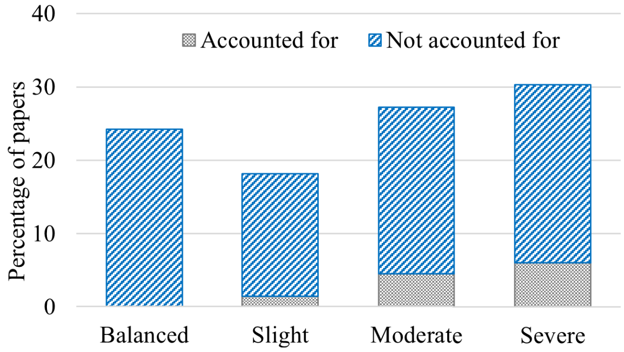

FIGURE 6. Breakdown of class imbalance (CI) observed within reviewed ML and agricultural literature.

that characterise agricultural data [56], [57]. Indeed, from the 100 papers in the cross-domain of ML & Agriculture studied in this survey the average dataset size was 797 [4]. CI is primarily determined based on the distribution of class representations within dataset [55]. The most common metric measures the frequency of instances within each class. Some techniques have been proposed to mitigate the problem of CI but these are not widely adopted in the agricultural domain. One such technique is the use of oversampling and undersampling [56], [58]. This technique works by manipulating the training data distribution to reduce imbalance. Oversampling duplicates the instances of the minority class. In contrast, undersampling eliminates instances from the majority class. Another technique is to impose higher cost on the learning process, such that models are penalised more heavily for misclassifying the minority class. In this way the learning parameters can be optimised to minimise the cost of misclassifying the model performance on both minority and majority classes. This is sometimes referred to as cost sensitive classification.

Sparsity of a dataset refers to its incompleteness [59], [60]. Sparsity is often measured by the number of zero attribute values in the dataset [59], [61]. For instance, yield data collected from different locations over time often have missing values for some for the features. A binary value for the presence or absence of frost is irrelevant for tropical regions. While ML models built on such data can give optimistic testing results, the model performs poorly when deployed [59], [62]. Such models can be characterised by unreliable weightings and high asymptotic errors [57], [60]. The common methods to mitigate the impacts are imputation and deletion. Imputation techniques generally apply to scenarios where the data sparsity is a result of missing value: that is, a situation where an attribute value is expected but none is available. An example of imputation technique is to replace the missing value with a mean or mode of the attribute value calculated from the other instances in the dataset. Deletion on the other hand removes attributes or instances from the dataset depending on the severity of the sparsity across instances or attributes respectively. Where possible, increasing the dataset size can increase the amount of information available. However, this approach is not always feasible within agricultural settings. Although the issue of sparsity is present in the agricultural setting, very few considerations are given to its potential impacts in the cross-domain of ML & Agriculture research.

High dimensionality (HD) occurs when the number of features within a dataset is similar or greater than the number of instances [63]–[65]. HD is measured by dividing the number of features (columns) by the number of instances (rows) [63], [65]. HD degrades the classifier quality as relationships between features and the target variable become more complex. This can lead to low classification accuracy and classifier performance on real world datasets [63]–[65]. Additionally, the number of features increases the complexity of the classifier which can reduce interpretability [64]. There are two main approaches to mitigate the impact of HD on classifiers. The first is to increase the dataset size, but this is not always feasible [65]. The second approach is to adopt feature engineering practices such as feature selection and transformation. Within the agricultural domain HD is more prevalent within hyperspectral datasets that are often utilised in remote sensing [66], [67].

3) APPLICATIONS OF ML (RQ2.3)

The widespread application of ML for the analysis of agricultural data provides insights into the research in various agricultural sectors. These applications include disease detection, soil classification and produce analysis. The reader is referred to Table 1 for the list of papers reviewed grouped by their sector and application.

The following details some of these specific applications seen in the literature. ML has been used to predict natural events such as famine [146], drought [42], [121] and frost [44], [120]. Indeed the work in [44] aimed to minimise the damage frost causes to crops through early prediction. By identifying the relationships between weather factors, a decision system was designed to make early binary predictions on frost events.

In [106] the authors presented a feasibility study on the use of novel e-nose for the detection of basal stem rot disease in oil palm. The various odours collected via the e-nose were analysed using NN to classify trees as healthy or infected. It was found that the novel system along with the integrated ML was able to recognise infected plants at a high rate of accuracy. Prior to this approach, the manual observations of visual signs were the predominant approach, but these signs present late in the disease life cycle leaving little time for treatment. Another work by [115] monitored the presence of a toxigenic fungus in maize. The authors compared the performances of logistic regression and DT, which performed similarly. Also identified were the features which contributed most to the contamination, this enabled improved management by farmers to reduce contamination risk. Other works such as [25], [40], [41], [107], [111], [116] also apply a variety of ML techniques for the detection of crop diseases.

[4]Note this number excludes spectral datasets which produce extreme outliers in this study.

TABLE 1. A summary of the reviewed ML and agricultural literature.

Agricultural sector	Application	ML algorithm	References
Soil	Soil type prediction	DT, BN, LogR, SVM	[48], [68], [69]
	Soil texture prediction	DT, RF, SVM, NN	[70], [71]
	Soil taxonomy prediction	kNN, DT, LogR, RF, SVM, NN, Other	[18], [19], [21], [72], [73]
	Soil properties prediction	DT, RF, SVM, Other	[43], [47], [71], [74], [75]
	Soil map production	Dt, RF, SVM, Other	[74], [76]–[78]
Produce analysis	Origins prediction	kNN, DT, RF, SVM, NN, Other	[16], [49], [52], [79]–[86]
	Cultivar / type classification	DT, LogR, SVM, NN, Other	[87]–[91]
	Ripeness / maturity classification	kNN, DT, RF, SVM, NN, Other	[89], [92]–[96]
	Quality prediction	kNN, BN, LogR, RF, SVM, NN	[80], [97]–[100]
	Nutrition prediction	RF	[17]
	Storage conditions prediction	NN	[101]
Yield	Yield prediction	kNN, DT, BN, RF, SVM, NN, Other	[20], [24], [51], [96], [102], [103]
	Yield contributor identification	DT, Other	[104]
	Harvest time prediction	RF	[105]
Pest Management	Disease prediction / detection	DT, BN, LogR, RF, SVM, NN, Other	[40], [41], [106]–[114]
	Fungus prediction / detection	DT, LogR, RF, NN	[25], [115]
	Pest prediction / detection	kNN, DT, BN, RF, SVM, NN, Other	[108], [116]–[118]
	Herbicide resistance prediction	DT	[119]
Natural events	Frost prediction	kNN, DT, BN, RF	[44], [120]
	Drought prediction	DT, RF, Other	[42], [121]
	Flood prediction	kNN, RF, SVM, NN, Other	[122]–[124]
	Landslide susceptibility prediction	DT, RF, Other	[125]
Other	Groundwater potential map production	DT, BN, LogR, RF, SVM, Other	[126]–[130]
	Plant stress identification	DT, RF, NN	[131], [132]
	Land cover classification	LogR, RF, SVM, NN, Other	[133]–[135]
	Crop type classification	kNN, RF, SVM, NN	[23], [136]–[139]
	Land suitability classification	DT, RF, SVM, NN	[3], [140], [141]
	Irrigation assessment	DT, Other	[13], [142], [143]
	Stream flow modelling	DT, NN, Other	[144]
	Heterotic classification	BN, SVM, Other	[145]

ML is often used in the soil sector on a variety of problems including the classification of soil type and class. In one study, [48] investigated the suitability of SVM to classify soil type and properties based on the chemical and physical features of the soil. The current methods of collecting physical and chemical data are laborious and time consuming. This success in applying SVM has high economic value. The authors of [19] compared DT, RF, kNN, NN and SVM to identify the best classifiers for the prediction of soil taxonomic groups. The authors found that kNN and SVM had the highest accuracy. Further works that apply ML to the soil sector include [18], [21], [43], [47].

ML classification has also been utilised in various aspects of crop yields including the forecasting of overall volume, marketable yields, yield variability and maturity of produce [20], [24], [51], [102], [103]. Work by [103] used SVM to develop a predictive models for rice yield and protein content. These models were used by farmers in management decision making to regulate growth conditions to achieve yield and protein content targets. The works in [89], [92], [93] also compared a variety of ML techniques for classifying the maturity of fruit, which helps with storage logistics and export of produce.

C. EVALUATION OF ML MODEL PERFORMANCE (RQ3)

Effective model assessment is key to the selection of the best ML classifier. The selection of the best method for model assessment should be directed by the objective of the task [34], [147]. That is, if the most accurate model is the objective than classification accuracy is appropriate. However, if the objective is to understand why a prediction was made then model interpretability is of greater importance. The most commonly used method to evaluate performance is classification accuracy. There have been recent moves to also account for model interpretability. While there has been utilisation within other domains, the literature review observed little adoption from the agricultural domain.

1) METHODS TO EVALUATE PERFORMANCE (RQ3.1)

The most commonly used method to evaluate performance as seen in the reviewed literature is classification accuracy. Within the reviewed literature, all the papers use some form of classification accuracy as a metric in the model evaluation. To the best of our knowledge the papers reviewed did not involve interpretability within their evaluation criteria. However, as will be discussed interpretability is critical when consequential decisions are based on model outputs [148]. As such interpretability for decision support systems has warranted its discussion in this work.

2) CLASSIFICATION ACCURACY (RQ3.2)

Classification accuracy is the measure of how well a model performs in correctly labelling a data instance. There are several metrics used to measure this, such as overall accuracy (OA), recall, and precision. The most commonly used metric observed within the literature is OA by over 70% of papers.

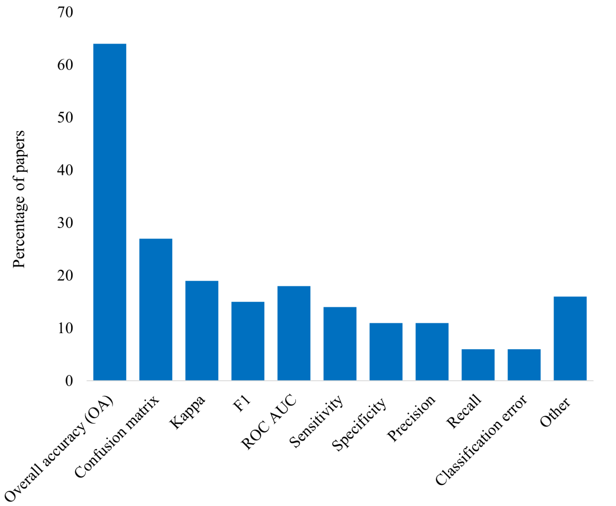

FIGURE 7. Performance metrics used in the reviewed ML and agricultural literature.

TABLE 2. Description of the datasets. Note: I - number of instances, C - number of classes, F - number of features, M - percent missing data.

Dataset	Data issue	I	C	F	M
Abalone	Severe CI	4177	28	8	0
Seeds	Balanced	210	3	7	0
Soybean	Mod CI & Mod sparsity	307	19	35	6

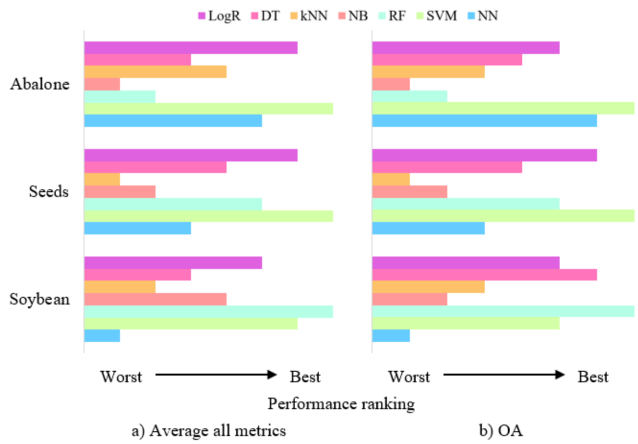

FIGURE 8. Classifier performance rankings based on datasets described in Table 2.

TABLE 3. Performance rankings for *Abalone* dataset described in Table 2 (Best - 1, Worst - 7).

Metric	LogR	DT	kNN	BN	RF	SVM	NN
OA	3	4	7	5	6	1	2
Kappa	3	4	7	5	6	1	2
F1	5	6	7	3	4	2	1
Precision	2	7	6	4	5	1	3
Recall	7	1	4	5	6	3	2

Figure 7 provides a breakdown of the metrics used within the 100 papers. OA is the percentage of correct predictions with respect to the total number of predictions. It is often chosen due to its simplicity and ease of interpretation [35]. While the use of this metric is appropriate for balanced datasets due to lower risk of learning bias, it can be inappropriate when CI is prevalent. Indeed, further analysis of OA found 44% of papers using OA presented with some level of CI, which may impact the validity of the reported accuracies. It is conceivable that the use of other metrics such as sensitivity, specificity, kappa, and F score are more appropriate. For contextualisation, the papers using OA with CI present were further analysed by looking at the datasets and paper objectives to identify whether there could have been other suitable alternative metrics. For example, it has been suggested in [41], and widely accepted in [17], [25], [120] that binary classification tasks where there is CI in the dataset, sensitivity and specificity are preferable models.

To further illustrate this, we conducted an experiment to compare classifier performance rankings based on a variety of performance metrics. The datasets used were obtained from the UCI repository [149] namely *Abalone, Seeds and Soybean* (Table 2). For all datasets, the data was split into 70% for training and 30% for testing, and six ML algorithms run. The algorithms included were: logistic regression (LogR), DT, kNN, naive bayes (NB), RF, SVM, and NN. For each algorithm hyper parameter selection occurred using GridSearchCV with k-fold cross validation (k=10). The test data were used to obtain the metrics OA, kappa, F1, precision, and recall.

The results showed that for each dataset, the ML classifier rankings differed. Figure 8 ranked the performance of the classifiers from best to worst, using an average of all five metrics, and OA. The first dataset *Abalone*, with severe CI, reported SVM as the best and NB as the worst technique. However, whether OA or an average were used changed the ranking of the other techniques. This illustrates the impact the selection of a metric has on comparing ML techniques. The second dataset, *Seeds* was balanced and resulted in the rankings of techniques remaining the same for the two performance metrics. This emphasises that OA can be an appropriate metric when the dataset is balanced. The third dataset, *Soybean* presenting with moderate CI and sparsity, returned SVM as the best technique and NN the worst, while the other techniques changed rankings between the two performance metrics. Review of the *Abalone* classifier rankings found that rankings varied with the different metrics (Table 3). It is worth pointing out that, the original objective for the generation of the *Abalone* dataset was to build a classifier which can predict the number of rings on the shell without the need to cut open and stain the shell. This process of manual counting is time-consuming, prone to error and laborious. The ability to know the age (i.e. young or old) is important to breeders who keep the young abalones for breeding while harvesting the older ones. The selection of the best model is dependent on which metric is used to assess performance. The best metric to select is dependent on the aim of the model. If the aim of the model is to identify the abalone to keep for breeding, i.e. the young, then it is important to identify all breeding stock. In this case recall would be preferred, which ranks DT as the best model. However, if the aim is to identify the abalone to harvest, i.e. the old, it is important to minimise false positives which would diminish the breading stock. The best metric in this case is precision which ranks SVM as the best model. Further, if the aim is to have a balance between recall and precision than F1 is the best metric to use which ranked NN as the highest.

3) MODEL INTERPRETABILITY (RQ3.3)

Model interpretability within the agricultural domain is rarely considered, in fact within the review of ML & Agriculture carried out in Section 3, no papers were found to measure interpretability[5]. However, understanding ML models is critical when consequential decisions are made based on the model output [148]. The first step in understanding a prediction is to establish the relationships between input features and the target variable. Interestingly, understanding of a model enables users to trust the model. This trust is essential for effective uptake and application of ML driven practices [147], [150], [151]. Within the agricultural domain this enables the use of ML driven precision agriculture practices within the farming community. There are two main types called interpretability and explainability which are expanded upon below [152]–[155].

Model interpretability has various levels of definitions between domains and applications [147]. However, a common theme is the ability for the human user to understand and engender trust [152], [156]–[162]. The inherent interpretability associated with interpretable models stems from the relatively simple structures that are used to represent knowledge within data. However, the simplicity in these structures mean that highly complex patterns in the data may be missed. This primarily makes such models less effective in discovering complex patterns and thus produce relatively lower accuracies compared to their non-interpretable counterparts. On the other hand, explainability refers to the post-hoc approach to confer information to the user rather than precisely how a model works [147], [155]. Used to primarily address the non-interpretable ML models, there are two common types of explainable models. The first model aims to render specific models understandable, while the second is model agnostic and aims to render any model understandable. Common techniques include LIME, COVAR and Anchor [163], [164]. The literature according to [34], [39], [155] recommend the use of interpretable models over non-interpretable models, which require additional models, where possible. For this reason we focus on interpretable models.

[5]However, [18] assigned models into groups based on their interpretability. While [91], [139] commented on the complexity of ML models.

TABLE 4. Performance of decision trees (Figure 9) using the *Seeds* dataset described in Table 2.

Dataset	Number of features	OA	Size	Time
All features	7	0.9206	9	0.475
Selected features	3	0.9206	6	0.475

Due to the range in definitions, requirements and applications of interpretability there is no widely applied metric to allow for measuring and comparing model interpretability. Throughout the wider computer science domain there are two metrics used to infer interpretability, often applied on rule based models such as DT. The first metric is a measure of model complexity developed by [165]. There are three parts to this metric, the first looks at the number of rules. As the number of rules increase, the likelihood of conflicting rules and increased difficulties in understanding all rules occurs. The second part measures the conditions per rule to account for the complexity and length of the rules. Thirdly, the conditions per classifier measure the overall complexity of the classifier regardless of the rules [165], [166]. The second metric is model size. This is a measure of the number of rules within a model. As model size increases, interpretability decreases [167].

To further illustrate the importance of considering interpretability an experiment was conducted. Using the *Seeds* dataset (Table 2), two datasets were created, the first dataset had all 7 features, while the second dataset used the top 3 features. These features were selected using feature importance rankings [168]. The datasets were split into 70% for training and 30% for testing and using GridSearchCV and 10-fold cross validation. A DT for each dataset was produced with the OA of both trees as 0.92 (Table 4). However, review of the tree structures (Figure 9) found large differences. The dataset with 7 features had a depth of five with nine rules (Figure 9.a), while the dataset with 3 features had a depth of three and six rules (Figure 9.b). If classification accuracy in the form of OA was used to select a model, either of the models would be suitable. However, if a more comprehensive assessment included model interpretability in the form of model size, the smaller model with three features would be selected due to higher implied interpretability.

IV. OPEN CHALLENGES IN ML & AGRICULTURE RESEARCH

In this section we discuss the challenges identified within this research and propose some directions of future research. Open and concerning challenge within the ML & Agriculture literature include the methods used to assess model performance, interdisciplinary research, application of IoT, and cyber security.

A. MODEL PERFORMANCE

There is scope to suggest that both classification accuracy and model interpretability should be considered in

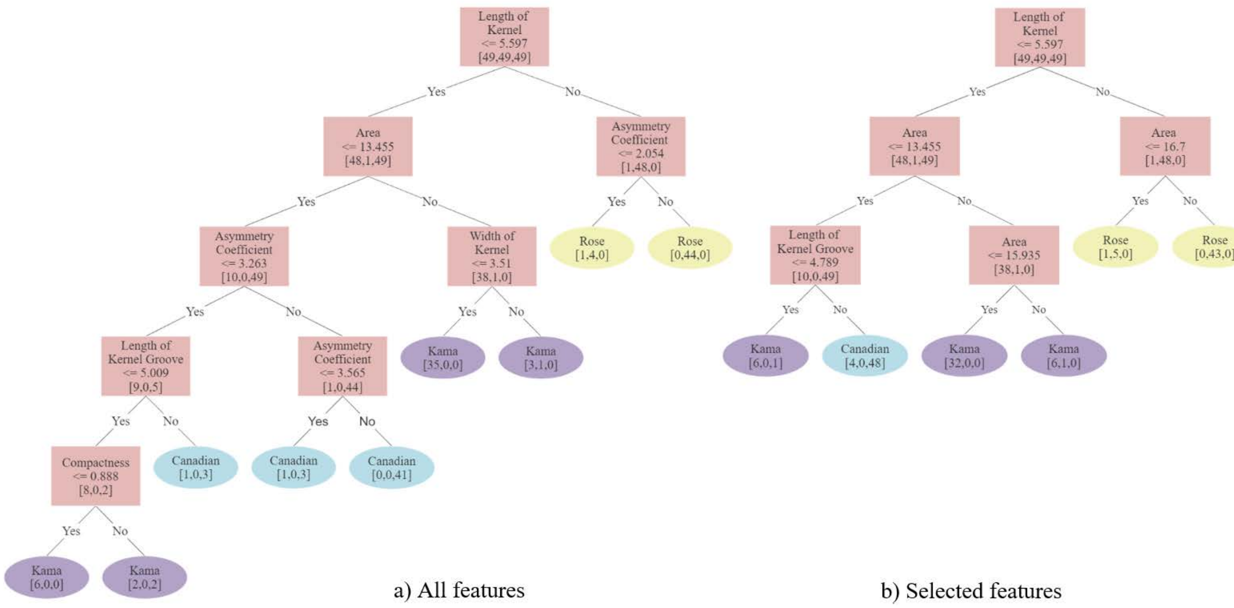

FIGURE 9. Decision trees detailed in Table 4 with the same overall accuracy, using the *Seeds* dataset described in Table 2.

assessing performance. It was found within the literature that there is a vast variation in the metrics used to assess classification accuracy while interpretability was rarely measured or accounted for. A possible future research path is the development of an assessment framework for benchmarking classifier performance. The recent works by [169] benchmarked the performance of models within the medical domain. Focus was on evaluating interpretability, fairness. Other recent works which benchmark interpretability focus on time series [170], images [171] and topic modelling [172]. Further to benchmarking within this domain, there is a clear need for interpretability in the cross-domain of ML & Agriculture. An interesting direction is how interpretability may play centre stage in the applications of ML into agriculture. In particular, can novel software tools which provide evidential support to the predictions of decision support systems be developed? A challenge that will need to be addressed is how to define a measure for interpretability within an agricultural setting; as we have already seen, different application scenarios may have different tenets with regards to interpretability. However, we believe this is plausible given the advances of ML research in developing meta learners for explaining models.

B. INTERDISCIPLINARY RESEARCH

Interdisciplinary research which merges various domain expertise with agriculture has resulted in rapid changes and improved agricultural production. Recent examples of successful interdisciplinary research include the scientific domains of chemistry and biology which saw the utilisation of nitrogen fertilisers and improved genetics. The current application of ML with agriculture is a feature of the interdisciplinary research of computer science and agriculture. However, the full benefits of this new research are yet to be fully realised as it experiences some fundamental barriers between domains. A major barrier is a disconnect between domain experts, limiting the effective uptake of new methodologies from the ML domain. Review of the literature found a large variation in methodologies and methods reported in agricultural research. Of note, only 44% of papers included any mention about the methods of pre-processing, and 47% on model optimisation techniques, while model trustworthiness was rarely mentioned. [31] concluded that the steep learning curve and lack of tools inhibit the uptake of ML techniques by agricultural researchers. [173] reported the major roadblock to ML application to be the lack of a fundamental understanding on ML, and [1], [32] found that many ML approaches were not well connected to the decision making processes. We have found that the adoption of best practices from the ML research domain is a key challenge in the utilisation of ML within the agricultural domain.

C. IoT CHALLENGES IN PRECISION AGRICULTURE

For a truly successful implementation of IoT in the agricultural domain, several challenges need to be considered and solved. While the major concern is interoperability there are three main types of limitations, which are at the application, network and device levels. Interoperability occurs at each of these levels within different dimensions including technical, syntactical, semantic and organisational. Limitations at the application level include how the data is analysed, however, due to the volume, diversity, and quality of the data, deriving worth from the data can be challenging. The use of ML for this analysis is rapidly increasing. Further, the data quality and availability needs to be addressed. Poor quality limits understanding of the data and in some cases compromise the success of IoT deployment.

At the network level, limitations include latency, in cloud or edge computing. However, 5G has the potential to improve peak throughput and increase data rates [7]. Further, communication range is of great concern in the setup of IoT systems especially in large scale applications. Selection of wireless or mesh technologies could improve this range. Network size and management need consideration in terms of the number of devices a network can handle and how they are deployed [7]. At the device level, IoT devices are constrained by power consumption and the hardiness of the device. Power consumption solutions can include energy harvesting, power efficient management and low power consumption sensors. IoT devices are generally located in harsh environments where conditions can damage the internal workings, this can lead to incorrect measurements, require unfeasible maintenance and re-calibration. General challenges to the successful application of IoT include a lack of products in some agricultural sectors, the complexity of the system in terms of the hardware and software, and the multifaceted nature of the processes being monitored. The scalability and flexibility for IoT to expand into various agricultural sectors also needs consideration, while the diverse data sources require standardisation to improve syntactical and semantic interoperability [7], [8], [174].

D. CYBER SECURITY IN PRECISION AGRICULTURE

Cyber security is the protection of information from attacks which impact confidentiality, integrity, availability and reliability of data [175]. These cyber threats may impact people, process, and/or technology, and originates from various actors. Actors include foreign governments, organised crime syndicates, insiders, and issue motivated groups [175]. Recent cyber attacks within the agricultural domain include the attack of Australian wool sales in 2020. Talman software which supplies over 75% of the Australian wool industry, fell victim to ransomware. This resulted in the shutdown of the software which services the AUS$ 60-80 million per week industry [175]. Another recent attack was against JBS Foods a global meat processing company in 2021. The attack resulted in the halting of operations ending once a ransom paid [176].

Cyber attacks can target various areas within precision agriculture, including the ML model and various IoT technologies. The ML models can be vulnerable to attacks. Namely, adversarial machine learning can decrease integrity of the ML models. These attacks often occur during the learning phase and can lead to missclassifications [177], [178]. Further, the dynamic and complex nature of IoT can lead to vulnerabilities [179]. A threat commonly observed with the use of IoT in precision agriculture include denial of service (DoS) attacks. DoS attacks can exhaust limited IoT device resources such as network capacity. Attacks can occur at the perception layer and may take the form of jamming, battery exhaustion, and collusion [180]. Another threat is ransomware, this is where attackers encrypt data rendering them unreadable until a ransom is met. Ransomware targets irreplaceable resources which increase willingness of the victim to pay. To be successful, the victim is prevented access [175], [181].

Cyber threats and vulnerabilities within the agricultural domain needs to be understood and accounted for in effective security strategies including proactive monitoring to deter and prevent threats. A key strategy is to ensure that all stake holders are aware and trained in mitigating common threats, and help to improve resilience [175]. Further, solutions to prevent unauthorised access, often seen with ransomware attacks, include restriction of access commonly known as effective access control [175], [181]. Protection of attacks on the IoT technology need to consider scalability of the IoT network and heterogeneity of IoT devices. Further, the solutions should cover multiple layers of the IoT infrastructure [180]. Finally, issues around privacy in data sharing for machine learning needs to be studied further. This is particularly challenging within the agriculture domain due to the complex supply chains associated with farms. It is conceivable that existing privacy preserving mechanisms such as [182] will be relevant in such applications.

V. CONCLUSION

In this work, we investigated the research within the cross-domain of machine learning and agriculture. We analysed the trends within the ML & Agriculture literature with a focus on the publications and thematic trends. Next, we discussed how ML is applied within the agricultural domain and identified some key agricultural data challenges which inhibit the effective application of common ML techniques. Also, we reviewed some common applications of ML and the agricultural problems they addressed. Further, we explored the methods of effective model assessment, and identified various pitfalls unbeknown to a large proportion of the literature. In doing so, we proposed a paradigm shift towards the use of alternative assessment metrics such as interpretability as a means of evaluating ML models. Finally, we presented some open challenges, key amongst these being the development of a unified framework for evaluating ML models within the agricultural domain; adoption of best practices from the ML research domain for a better utilisation of ML techniques; addressing the limitations associated with the implementation of IoT technologies; and understanding and mitigating the cyber security threat landscape within precision agriculture.

REFERENCES

[1] K. G. Liakos, P. Busato, D. Moshou, and S. Pearson, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, p. 2674, 2018, doi: 10.3390/s18082674.

[2] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," *IEEE Access*, vol. 9, pp. 4843–4873, 2021.

[3] R. Taghizadeh-Mehrjardi, K. Nabiollahi, L. Rasoli, R. Kerry, and T. Scholten, "Land suitability assessment and agricultural production sustainability using machine learning models," *Agronomy*, vol. 10, no. 4, p. 573, Apr. 2020, doi: 10.3390/agronomy10040573.

[4] M. Keogh and M. Henry, "The implications of digital agriculture and big data for Australian agriculture," Austral. Farm Inst., Sydney, NSW, Australia, Tech. Rep., 2016.

[5] S. Ahmed, "Security and privacy in smart cities: Challenges and opportunities," *Int. J. Eng. Trends Technol.*, vol. 68, no. 2, pp. 1–8, Feb. 2020, doi: 10.14445/22315381/IJETT-V68I2P201.

[6] J. M. Talavera, L. E. Tobón, J. A. Gómez, M. A. Culman, J. M. Aranda, D. T. Parra, L. A. Quiroz, A. Hoyos, and L. E. Garreta, "Review of IoT applications in agro-industrial and environmental fields," *Comput. Electron. Agricult.*, vol. 142, pp. 283–297, Nov. 2017, doi: 10.1016/j.compag.2017.09.015.

[7] 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," *Biosyst. Eng.*, vol. 191, pp. 60–84, Mar. 2020, doi: 10.1016/j.biosystemseng.2019.12.013.

[8] G. Vitali, M. Francia, M. Golfarelli, and M. Canavari, "Crop management with the IoT: An interdisciplinary survey," *Agronomy*, vol. 11, no. 1, p. 181, Jan. 2021, doi: 10.3390/agronomy11010181.

[9] A. T. M. Shakil Ahamed, N. T. Mahmood, N. Hossain, M. T. Kabir, K. Das, F. Rahman, and R. M. Rahman, "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh," in *Proc. IEEE/ACIS 16th Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput. (SNPD)*, Jun. 2015, pp. 1–6, doi: 10.1109/SNPD.2015.7176185.

[10] 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," *IEEE Access*, vol. 7, pp. 156237–156271, 2019, doi: 10.1109/ACCESS.2019.2949703.

[11] O. Green, E. S. Nadimi, V. Blanes-Vidal, R. N. Jørgensen, I. M. L. D. Storm, and C. G. Sørensen, "Monitoring and modeling temperature variations inside silage stacks using novel wireless sensor networks," *Comput. Electron. Agricult.*, vol. 69, no. 2, pp. 149–157, Dec. 2009, doi: 10.1016/j.compag.2009.07.021.

[12] K. Langendoen, A. Baggio, and O. Visser, "Murphy loves potatoes: Experiences from a pilot sensor network deployment in precision agriculture," in *Proc. 20th IEEE Int. Parallel Distrib. Process. Symp. (IPDPS)*, Apr. 2006, p. 8, doi: 10.1109/IPDPS.2006.1639412.

[13] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, G. Ravid, "Applying machine learning on sensor data for irrigation recommendations: Revealing the agronomist's tacit knowledge," *Precis. Agricult.*, vol. 19, no. 3, pp. 421–444, 2018, doi: 10.1007/s11119-017-9527-4.

[14] C. A. Gonzalez-Amarillo, J. C. Corrales-Munoz, M. A. Mendoza-Moreno, and A. M. Gonzalez Amarillo, A. F. Hussein, N. Arunkumar, and G. Ramirez-Gonzalez, "An IoT-based traceability system for greenhouse seedling crops," *IEEE Access*, vol. 6, pp. 67528–67535, 2018, doi: 10.1109/ACCESS.2018.2877293.

[15] Z. Pang, Q. Chen, W. Han, and L. Zheng, "Value-centric design of the Internet-of-things solution for food supply chain: Value creation, sensor portfolio and information fusion," *Inf. Syst. Frontiers*, vol. 17, no. 2, pp. 289–319, Apr. 2015, doi: 10.1007/s10796-012-9374-9.

[16] O. Anjos, C. Iglesias, F. Peres, and J. Martínez, A. García, and J. Taboada, "Neural networks applied to discriminate botanical origin of honeys," *Food Chem.*, vol. 175, pp. 128–136, May 2015, doi: 10.1016/j.foodchem.2014.11.121.

[17] D. L. Betemps and B. V. De Paula, S.-É. Parent, S. P. Galarça, N. A. Mayer, G. A. B. Marodin, D. E. Rozane, W. Natale, G. W. B. Melo, L. E. Parent, and G. Brunetto, "Humboldtian diagnosis of peach tree *(Prunus persica)* nutrition using machine-learning and compositional methods," *Agronomy*, vol. 10, no. 6, p. 900, 2020, doi: 10.3390/agronomy10060900.

[18] C. W. Brungard, J. L. Boettinger, M. C. Duniway, S. A. Wills, and T. C. Edwards, "Machine learning for predicting soil classes in three semi-arid landscapes," *Geoderma*, vols. 239–240, pp. 68–83, Feb. 2015, doi: 10.1016/j.geoderma.2014.09.019.

[19] B. Heung, H. C. Ho, J. Zhang, A. Knudby, C. E. Bulmer, and M. G. Schmidt, "An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping," *Geoderma*, vol. 265, pp. 62–77, Mar. 2016, doi: 10.1016/j.geoderma.2015.11.014.

[20] J. R. Romero, P. F. Roncallo, P. C. Akkiraju, I. Ponzoni, V. C. Echenique, and J. A. Carballido, "Using classification algorithms for predicting durum wheat yield in the province of buenos aires," *Comput. Electron. Agricult.*, vol. 96, pp. 173–179, Aug. 2013, doi: 10.1016/j.compag.2013.05.006.

[21] M. Lacoste, B. Lemercier, and C. Walter, "Regional mapping of soil parent material by machine learning based on point data," *Geomorphology*, vol. 133, nos. 1–2, pp. 90–99, Oct. 2011, doi: 10.1016/j.geomorph.2011.06.026.

[22] N. Mansuy, E. Thiffault, and D. Paré, P. Bernier, L. Guindon, P. Villemaire, V. Poirier, and A. Beaudoin, "Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the K-nearest neighbor method," *Geoderma*, vols. 235–236, pp. 59–73, Dec. 2014, doi: 10.1016/j.geoderma.2014.06.032.

[23] M. Trebar and N. Steele, "Application of distributed SVM architectures in classifying forest data cover types," *Comput. Electron. Agricult.*, vol. 63, no. 2, pp. 119–130, Oct. 2008, doi: 10.1016/j.compag.2008.02.001.

[24] A. Chemchem, F. Alin, and M. Krajecki, "Combining SMOTE sampling and machine learning for forecasting wheat yields in France," in *Proc. IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. (AIKE)*, Jun. 2019, pp. 9–14, doi: 10.1109/AIKE.2019.00010.

[25] L. K. Mehra, C. Cowger, K. Gross, and P. S. Ojiambo, "Predicting preplanting risk of stagonospora nodorum blotch in winter wheat using machine learning models," *Frontiers Plant Sci.*, vol. 7, pp. 390–404, Mar. 2016, doi: 10.3389/fpls.2016.00390.

[26] D. Elavarasan, D. R. Vinvent, V. Sharma, A. Y. Zomaya, and K. Srinivasan, "Forecasting yield by integrating agrarian factors and machine learning models: A survey," *Comput. Electron. Agricult.*, vol. 155, pp. 257–282, Dec. 2018, doi: 10.1016/j.compag.2018.10.024.

[27] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Comput. Electron. Agricult.*, vol. 151, pp. 61–69, Aug. 2018, doi: 10.1016/j.compag.2018.05.012.

[28] T. V. Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105709, doi: 10.1016/j.compag.2020.105709.

[29] C. Maione, F. Barbosa, and R. M. Barbosa, "Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review," *Comput. Electron. Agricult.*, vol. 157, pp. 436–446, Feb. 2019, doi: 10.1016/j.compag.2019.01.020.

[30] S. Raghavendra. N and P. C. Deka, "Support vector machine applications in the field of hydrology: A review," *Appl. Soft Comput.*, vol. 19, pp. 372–386, Jun. 2014, doi: 10.1016/j.asoc.2014.02.002.

[31] B. Drury, J. Valverde-Rebaza, M. F. Moura, and A. de Andrade Lopes, "A survey of the applications of Bayesian networks in agriculture," *Eng. Appl. Artif. Intell.*, vol. 65, pp. 29–42, Oct. 2017, doi: 10.1016/j.engappai.2017.07.003.

[32] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," *Sensors*, vol. 21, no. 11, p. 3758, 2021, doi: 10.3390/s21113758.

[33] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.

[34] J. Padarian, B. Minasny, and A. B. McBratney, "Machine learning and soil sciences: A review aided by machine learning tools," *Soil*, vol. 6, no. 1, pp. 35–52, Feb. 2020, doi: 10.5194/soil-6-35-2020.

[35] J. Lever, M. Krzywinski, and N. Altman, "Points of Significance: Classification evaluation," *Nature Methods*, vol. 13, no. 8, pp. 603–604, 2016, doi: 10.1038/nmeth.3945.

[36] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, and H. Yuanyue, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017, doi: 10.1016/j.eswa.2016.12.035.

[37] F. Provast, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," in *Proc. 15th Int. Conf. Mach. Learn.* San Mateo, CA, USA: Morgan Kaufmann, 1997, pp. 445–453.

[38] S. Saia, N. Nelson, A. Huseth, K. Grieger, and B. Reich, "Transitioning machine learning from theory to practice in natural resources management," in *Ecological Modelling*, vol. 435. 2020, doi: 10.1016/j.ecolmodel.2020.109257.

[39] O. Loyola-Gonzalez, "Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view," *IEEE Access*, vol. 7, pp. 154096–154113, 2019, doi: 10.1109/ACCESS.2019.2949286.

[40] M. Ilic, S. Ilic, S. Jovic, and S. Panic, "Early cherry fruit pathogen disease detection based on data mining prediction," *Comput. Electron. Agricult.*, vol. 150, pp. 418–425, Jul. 2018, doi: 10.1016/j.compag.2018.05.008.

[41] K. S. Kim, R. M. Beresford, and M. Walter, "Development of a disease risk prediction model for downy mildew *(Peronospora sparsa)* in boysenberry," *Phytopathology*, vol. 104, no. 1, pp. 50–56, Jan. 2014, doi: 10.1094/PHYTO-02-13-0058-R.

[42] J. Rhee and J. Im, "Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data," *Agricult. Forest Meteorol.*, vols. 237–238, pp. 105–122, May 2017, doi: 10.1016/j.agrformet.2017.02.011.

[43] S. Rajeswari and K. Suthendran, "C5.0: Advanced decision tree (ADT) classification model for agricultural data analysis on cloud," *Comput. Electron. Agricult.*, vol. 156, pp. 530–539, Jan. 2019, doi: 10.1016/j.compag.2018.12.013.

[44] J. M. Cadenas, M. C. Garrido, R. Martínez-España, and M. A. Guillén-Navarro, "Making decisions for frost prediction in agricultural crops in a soft computing framework," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105587, doi: 10.1016/j.compag.2020.105587.

[45] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, 2018, doi: 10.1145/3236009.

[46] W. S. Noble, "What is a support vector machine?" *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006, doi: 10.1038/nbt1206-1565.

[47] Y. Liu, H. Wang, H. Zhang, and K. Liber, "A comprehensive support vector machine-based classification model for soil quality assessment," *Soil Tillage Res.*, vol. 155, pp. 19–26, Jan. 2016, doi: 10.1016/j.still.2015.07.006.

[48] M. Kovačević, B. Bajat, and B. Gajić, "Soil type classification and estimation of soil properties using support vector machines," *Geoderma*, vol. 15, nos. 3–4, pp. 340–347, 2010, doi: 10.1016/j.geoderma.2009.11.005.

[49] C. Maione, B. L. Batista, A. D. Campiglia, F. Barbosa, and R. M. Barbosa, "Classification of geographic origin of Rice by data mining and inductively coupled plasma mass spectrometry," *Comput. Electron. Agricult.*, vol. 121, pp. 101–107, Feb. 2016, doi: 10.1016/j.compag.2015.11.009.

[50] Y. Huang, Y. Lan, S. J. Thomson, F. Alex, W. C. Hoffmann, and R. E. Lacey, "Development of soft computing and applications in agricultural and biological engineering," *Comput. Electron. Agricult.*, vol. 71, pp. 107–127, May 2010.

[51] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agricult.*, vol. 121, pp. 57–65, Feb. 2016, doi: 10.1016/j.compag.2015.11.018.

[52] L. Chen, J. Wang, Z. Ye, J. Zhao, X. Xue, Y. V. Heyden, and Q. Sun, "Classification of Chinese honeys according to their floral origin by near infrared spectroscopy," *Food Chem.*, vol. 135, no. 2, pp. 338–342, Nov. 2012, doi: 10.1016/j.foodchem.2012.02.156.

[53] J. Zhang and I. Mani, "kNN approach to unbalanced data distributions: A case study involving information extraction," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Washington, DC, USA, 2003, pp. 1–7.

[54] C. D'Este, G. Timms, A. Turnbull, and A. Rahman, "Ensemble aggregation methods for relocating models of rare events," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 58–65, Sep. 2014, doi: 10.1016/j.engappai.2014.05.007.

[55] J. Ortigosa-Hernández, I. Inza, and J. A. Lozano, "Measuring the class-imbalance extent of multi-class problems," *Pattern Recognit. Lett.*, vol. 98, pp. 32–38, Oct. 2017, doi: 10.1016/j.patrec.2017.08.002.

[56] N. Japkpwocz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, 2000, pp. 429–449, doi: 10.3233/IDA-2002-6504.

[57] R. Whitsed, R. Corner, and S. Cook, "A model to predict ordinal suitability using sparse and uncertain data," *Appl. Geography*, vol. 32, no. 2, pp. 401–408, Mar. 2012. http://researchoutput.csu.edu.au/R/-?func=dbin-jump-full&object_id=29949&local_base=GEN01-CSU01, doi: 10.1016/j.apgeog.2011.06.016.

[58] G. M. Weiss, "Mining with rarity: A unifying framework," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 7–19, 2004.

[59] O. Wärnling and J. Bissmark, "The sparse data problem within classification algorithms: The effect of sparse data on the Naïve Bayes algorithm," Ph.D. thesis, CSC, KTH Roy. Inst. Technol., Sweden, 2017. [Online]. Available: http://www.diva-portal.se/smash/get/diva2:1111045/FULLTEXT01.pdf

[60] W. Zhang, X. Zhang, H. Wang, and D. Chen, "A deep variational matrix factorization method for recommendation on large scale sparse dataset," *Neurocomputing*, vol. 334, pp. 206–218, Mar. 2019, doi: 10.1016/j.neucom.2019.01.028.

[61] S. Greenland, M. A. Mansournia, and D. G. Altman, "Sparse data bias: A problem hiding in plain sight," *Bmj*, vol. 352, no. 1981, pp. 1–6, 2016, doi: 10.1136/bmj.i1981.

[62] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009, doi: 10.1109/TIT.2009.2027527.

[63] S. X. Chen and Y.-L. Qin, "A two-sample test for high-dimensional data with applications to gene-set testing," *Ann. Statist.*, vol. 38, no. 2, pp. 808–835, 2010, doi: 10.1214/09-AOS716.

[64] C. R. da Silva, J. A. S. Centeno, and S. R. Aranha, "Reduction of the dimensionality of hyperspectral data for the classification of agricultural scenes," in *Proc. 13th FIG Symp. Deformation Meas. Anal.*, 2008, pp. 1–10.

[65] J. K. Gilbertson and A. van Niekerk, "Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning," *Comput. Electron. Agricult.*, vol. 142, pp. 50–58, Nov. 2017, doi: 10.1016/j.compag.2017.08.024.

[66] J. Hruška, T. Adão, L. Pádua, P. Marques, A. Cunha, E. Peres, A. Sousa, R. Morais, and J. J. Sousa, "Machine learning classification methods in hyperspectral data processing for agricultural applications," in *Proc. Int. Conf. Geoinform. Data Anal.*, Apr. 2018, pp. 137–141, doi: 10.1145/3220228.3220242.

[67] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018, doi: 10.1080/01431161.2018.1433343.

[68] P. Scull, J. Franklin, and O. A. Chadwick, "The application of classification tree analysis to soil type prediction in a desert landscape," *Ecol. Model.*, vol. 181, no. 1, pp. 1–15, Jan. 2005, doi: 10.1016/j.ecolmodel.2004.06.036.

[69] A. Jafari, P. A. Finke, and J. Vande Wauw, S. Ayoubi, and H. Khademi, "Spatial prediction of USDA- great soil groups in the arid Zarand region, Iran: Comparing logistic regression approaches to predict diagnostic horizons and soil types," *Eur. J. Soil Sci.*, vol. 63, no. 2, pp. 284–298, 2012, doi: 10.1111/j.1365-2389.2012.01425.x.

[70] W. Wu, A.-D. Li, X.-H. He, R. Ma, H.-B. Liu, and J.-K. Lv, "A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China," *Comput. Electron. Agricult.*, vol. 144, pp. 86–93, Jan. 2018, doi: 10.1016/j.compag.2017.11.037.

[71] R. Pittman, B. Hu, and K. Webster, "Improvement of soil property mapping in the great clay belt of northern Ontario using multi-source remotely sensed data," *Geoderma*, vol. 381, Jan. 2021, Art. no. 114761, doi: 10.1016/j.geoderma.2020.114761.

[72] A. Jafari, S. Ayoubi, H. Khademi, P. A. Finke, and N. Toomanian, "Selection of a taxonomic level for soil mapping using diversity and map purity indices: A case study from an Iranian arid region," *Geomorphology*, vol. 201, pp. 86–97, Nov. 2013, doi: 10.1016/j.geomorph.2013.06.010.

[73] T. Hengl, N. Toomanian, H. I. Reuter, and M. J. Malakouti, "Methods to interpolate soil categorical variables from profile observations: Lessons from Iran," *Geoderma*, vol. 140, no. 4, pp. 417–427, Aug. 2007, doi: 10.1016/j.geoderma.2007.04.022.

[74] V. Estévez, A. Beucher, S. Mattbäck, A. Boman, J. Auri, K.-M. Björk, and P. Österholm, "Machine learning techniques for acid sulfate soil mapping in southeastern Finland," *Geoderma*, vol. 406, Jan. 2022, Art. no. 115446, doi: 10.1016/j.geoderma.2021.115446.

[75] H. Wang, H. Zhang, and Y. Liu, "Using a posterior probability support vector machine model to assess soil quality in Taiyuan, China," *Soil Tillage Res.*, vol. 185, pp. 146–152, Jan. 2019, doi: 10.1016/j.still.2018.08.013.

[76] M. R. Pahlavan Rad, N. Toomanian, F. Khormali, C. W. Brungard, C. B. Komaki, and P. Bogaert, "Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran," *Geoderma*, vols. 232–234, pp. 97–106, Nov. 2014, doi: 10.1016/j.geoderma.2014.04.036.

[77] B. Heung, C. E. Bulmer, and M. G. Schmidt, "Predictive soil parent material mapping at a regional-scale: A random forest approach," *Geoderma*, vols. 214–215, pp. 141–154, Feb. 2014, doi: 10.1016/j.geoderma.2013.09.016.

[78] C. Grinand, D. Arrouays, B. Laroche, and M. P. Martin, "Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context," *Geoderma*, vol. 143, nos. 1–2, pp. 180–190, Jan. 2008, doi: 10.1016/j.geoderma.2007.11.004.

[79] B. L. Batista, L. R. S. da Silva, B. A. Rocha, J. L. Rodrigues, A. A. Berretta-Silva, T. O. Bonates, V. S. D. Gomes, R. M. Barbosa, and F. Barbosa, "Multi-element determination in Brazilian honey samples by inductively coupled plasma mass spectrometry and estimation of geographic origin with data mining techniques," *Food Res. Int.*, vol. 49, no. 1, pp. 209–215, Nov. 2012, doi: 10.1016/j.foodres.2012.07.015.

[80] B. V. Canizo, L. B. Escudero, R. G. Pellerano, and R. G. Wuilloud, "Data mining approach based on chemical composition of grape skin for quality evaluation and traceability prediction of grapes," *Comput. Electron. Agricult.*, vol. 162, pp. 514–522, Jul. 2019, doi: 10.1016/j.compag.2019.04.043.

[81] J. M. Gutiérrez, Z. Haddi, A. Amari, B. Bouchikhi, A. Mimendia, and X. Cetó, and M. Del Valle, "Hybrid electronic tongue based on multisensor data fusion for discrimination of beers," *Sens. Actuators B, Chem.*, vol. 177, pp. 989–996, Feb. 2013, doi: 10.1016/j.snb.2012.11.110.

[82] M. A. Rodríguez-Delgado, G. González-Hernández, J. E. Conde-González, and J. P. Pérez-Trujillo, "Principal component analysis of the polyphenol content in young red wines," *Food Chem.*, vol. 78, no. 4, pp. 523–532, 2002, doi: 10.1016/S0308-8146(02)00206-6.

[83] E. Marengo and M. Aceto, "Statistical investigation of the differences in the distribution of metals in Nebbiolo-based wines," *Food Chem.*, vol. 81, no. 4, pp. 621–630, 2003, doi: 10.1016/S0308-8146(02)00564-2.

[84] Y. Sun, N. Liu, X. Kang, Y. Zhao, R. Cao, J. Ning, H. Ding, X. Sheng, and D. Zhou, "Rapid identification of geographical origin of sea cucumbers apostichopus japonicus using FT-NIR coupled with light gradient boosting machine," *Food Control*, vol. 124, Jun. 2021, Art. no. 107883, doi: 10.1016/j.foodcont.2021.107883.

[85] S. Popek, M. Halagarda, and K. Kursa, "A new model to identify botanical origin of Polish honeys based on the physicochemical parameters and chemometric analysis," *LWT*, vol. 77, pp. 482–487, Apr. 2017, doi: 10.1016/j.lwt.2016.12.003.

[86] Z. Gan, Y. Yang, J. Li, X. Wen, M. Zhu, Y. Jiang, and Y. Ni, "Using sensor and spectral analysis to classify botanical origin and determine adulteration of raw honey," *J. Food Eng.*, vol. 178, pp. 151–158, Jun. 2016, doi: 10.1016/j.jfoodeng.2016.01.016.

[87] E. Schuhfried and J. Sánchez Del Pulgar, M. Bobba, R. Piro, L. Cappellin, T. D. Märk, and F. Biasioli, "Classification of 7 monofloral honey varieties by PTR-ToF-MS direct headspace analysis and chemometrics," *Talanta*, vol. 147, pp. 213–219, Jan. 2016, doi: 10.1016/j.talanta.2015.09.062.

[88] H. Zhao, Y. Zhan, Z. Xu, and J. John Nduwamungu, Y. Zhou, R. Powers, and C. Xu, "The application of machine-learning and Raman spectroscopy for the rapid detection of edible oils type and adulteration," *Food Chem.*, vol. 373, Mar. 2022, Art. no. 131471, doi: 10.1016/j.foodchem.2021.131471.

[89] S. Benedetti, S. Buratti, A. Spinardi, S. Mannino, and I. Mignani, "Electronic nose as a non-destructive tool to characterise peach cultivars and to monitor their ripening stage during shelf-life," *Postharvest Biol. Technol.*, vol. 47, no. 2, pp. 181–188, Feb. 2008, doi: 10.1016/j.postharvbio.2007.06.012.

[90] M. J. Oates, P. Fox, and L. Sanchez-Rodriguez, Á. A. Carbonell-Barrachina, and A. Ruiz-Canales, "DFT based classification of olive oil type using a sinusoidally heated, low cost electronic nose," *Comput. Electron. Agricult.*, vol. 155, pp. 348–358, Dec. 2018, doi: 10.1016/j.compag.2018.10.026.

[91] K. Brudzewski, "Classification of milk by means of an electronic nose and SVM neural network," *Sens. Actuators B, Chem.*, vol. 98, nos. 2–3, pp. 291–298, Mar. 2004, doi: 10.1016/j.snb.2003.10.028.

[92] J. A. Caladcad, S. Cabahug, M. R. Catamco, P. E. Villaceran, L. Cosgafa, K. N. Cabizares, M. Hermosilla, and E. J. Piedad, "Determining philippine coconut maturity level using machine learning algorithms based on acoustic signal," *Comput. Electron. Agricult.*, vol. 172, May 2020, Art. no. 105327, doi: 10.1016/j.compag.2020.105327.

[93] H. G. J. Voss, S. L. Stevan, and R. A. Ayub, "Peach growth cycle monitoring using an electronic nose," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104858, doi: 10.1016/j.compag.2019.104858.

[94] J. Brezmes, E. Llobet, X. Vilanova, G. Saiz, and X. Correig, "Fruit ripeness monitoring using an electronic nose," *Sens. Actuators B, Chem.*, vol. 69, no. 3, pp. 223–229, 2000, doi: 10.1016/S0925-4005(00)00494-9.

[95] D. Moshou, S. Wahlen, R. Strasser, A. Schenk, and H. Ramon, "Apple mealiness detection using fluorescence and self-organising maps," *Comput. Electron. Agricult.*, vol. 40, nos. 1–3, 2003, pp. 103–114, doi: 10.1016/S0168-1699(03)00014-0.

[96] Y.-X. Su, H. Xu, and L.-J. Yan, "Support vector machine-based open crop model (SBOCM): Case of Rice production in China," *Saudi J. Biol. Sci.*, vol. 24, no. 3, pp. 537–547, Mar. 2017, doi: 10.1016/j.sjbs.2017.01.024.

[97] C. Li, P. Heinemann, and R. Sherry, "Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection," *Sens. Actuators B, Chem.*, vol. 125, no. 1, pp. 301–310, Jul. 2007, doi: 10.1016/j.snb.2007.02.027.

[98] A. A. Argyri, E. Z. Panagou, P. A. Tarantilis, M. Polysiou, and G. J. Nychas, "Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks," *Sens. Actuators B, Chem.*, vol. 145, no. 1, pp. 146–154, 2010, doi: 10.1016/j.snb.2009.11.052.

[99] M. Bizzani, D. W. M. Flores, L. A. Colnago, and M. D. Ferreira, "Monitoring of soluble pectin content in orange juice by means of MIR and TD-NMR spectroscopy combined with machine learning," *Food Chem.*, vol. 332, Art. no. 127383, Dec. 2020, doi: 10.1016/j.foodchem.2020.127383.

[100] C. Bielza, P. Barreiro, M. I. Rodríguez-Galiano, and J. Martín, "Logistic regression for simulating damage occurrence on a fruit grading line," *Comput. Electron. Agricult.*, vol. 39, no. 2 pp. 95–113, 2003, doi: 10.1016/S0168-1699(03)00021-8.

[101] A. Riul and H. C. De Sousa, R. R. Malmegrim, D. S. Dos Santos, A. C. Carvalho, F. J. Fonseca, O. N. Oliveira, and L. H. C. Mattoso, "Wine classification by taste sensors made from ultra-thin films and using neural networks," *Sens. Actuators B, Chem.*, vol. 98, no. 1, pp. 77–82, 2004, doi: 10.1016/j.snb.2003.09.025.

[102] S. Fukuda, W. Spreer, E. Yasunaga, K. Yuge, V. Sardsud, and J. Müller, "Random Forests modelling for the estimation of mango *(Mangifera indica L. cv. Chok Anan)* fruit yields under different irrigation regimes," *Agricult. Water Manage.*, vol. 116, pp. 142–150, Jan. 2013, doi: 10.1016/j.agwat.2012.07.003.

[103] K. Saruta, Y. Hirai, K. Tanaka, E. Inoue, T. Okayasu, and M. Mitsuoka, "Predictive models for yield and protein content of Brown Rice using support vector machine," *Comput. Electron. Agricult.*, vol. 99, pp. 93–100, Nov. 2013, doi: 10.1016/j.compag.2013.09.003.

[104] A. Shekoofa, Y. Emam, N. Shekoufa, M. Ebrahimi, and E. Ebrahimie, "Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: A new avenue in intelligent agriculture," *PLoS ONE*, vol. 9, no. 5, May 2014, Art. no. e97288, doi: 10.1371/journal.pone.0097288.

[105] D. Yu, N. Lord, J. Polk, K. Dhakal, S. Li, Y. Yin, S. E. Duncan, H. Wang, B. Zhang, and H. Huang, "Physical and chemical properties of edamame during bean development and application of spectroscopy-based machine learning methods to predict optimal harvest time," *Food Chem.*, vol. 368, Jan. 2022, Art. no. 130799, doi: 10.1016/j.foodchem.2021.130799.

[106] M. A. Markom, A. Y. M. Shakaff, A. H. Adom, M. N. Ahmad, W. Hidayat, A. H. Abdullah, and N. A. Fikri, "Intelligent electronic nose system for basal stem rot disease detection," *Comput. Electron. Agricult.*, vol. 66, no. 2, pp. 140–146, May 2009, doi: 10.1016/j.compag.2009.01.006.

[107] K. M. Baker and W. W. Kirk, "Comparative analysis of models integrating synoptic forecast data into potato late blight risk estimate systems," *Comput. Electron. Agricult.*, vol. 57, no. 1, pp. 23–32, May 2007, doi: 10.1016/j.compag.2007.01.011.

[108] H. Ma, W. Huang, Y. Jing, C. Yang, L. Han, Y. Dong, H. Ye, Y. Shi, Q. Zheng, L. Liu, and C. Ruan, "Integrating growth and environmental parameters to discriminate powdery mildew and aphid of winter wheat using bi-temporal Landsat-8 imagery," *Remote Sens.*, vol. 11, no. 7, p. 846, 2019, doi: 10.3390/RS11070846.

[109] D. Moshou, C. Bravo, J. West, S. Wahlen, A. Mccartney, and H. Ramon, "Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks," *Comput. Electron. Agricult.*, vol. 44, pp. 173–188, Sep. 2004, doi: 10.1016/j.compag.2004.04.003.

[110] A. Chaudhary, S. Kolhe, and R. Kamal, "A hybrid ensemble for classification in multiclass datasets: An application to oilseed disease dataset," *Comput. Electron. Agricult.*, vol. 124, pp. 65–72, Jun. 2016, doi: 10.1016/j.compag.2016.03.026.

[111] E. D. De Wolf, L. V. Madden, and P. E. Lipps, "Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data," *Phytopathology*, vol. 93, no. 4, pp. 428–435, 2003, doi: 10.1094/PHYTO.2003.93.4.428.

[112] M. Chen, F. Brun, M. Raynal, and D. Makowski, "Forecasting severe grape downy mildew attacks using machine learning," *PLoS ONE*, vol. 15, no. 3, Mar. 2020, Art. no. e0230254, doi: 10.1371/journal.pone.0230254.

[113] C. E. Cabrera Ardila, L. Alberto Ramirez, and F. A. Prieto Ortiz, "Spectral analysis for the early detection of anthracnose in fruits of sugar mango (Mangifera indica)," *Comput. Electron. Agricult.*, vol. 173, Jun. 2020, Art. no. 105357, doi: 10.1016/j.compag.2020.105357.

[114] D. Moshou, C. Bravo, R. Oberti, J. West, L. Bodria, A. Mccartney, and H. Ramon, "Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using Kohonen maps," *Real-Time Imag.*, vol. 11, pp. 75–83, Apr. 2005, doi: 10.1016/j.rti.2005.03.003.

[115] P. Battilani, A. Pietri, C. Barbano, A. Scandolara, T. Bertuzzi, and A. Marocco, "Logistic regression modeling of cropping systems to predict fumonisin contamination in maize," *J. Agricult. Food Chem.*, vol. 56, no. 21, pp. 10433–10438, 2008, doi: 10.1021/jf801809d.

[116] M. G. Hill, P. G. Connolly, P. Reutemann, and D. Fletcher, "The use of data mining to assist crop protection decisions on kiwifruit in New Zealand," *Comput. Electron. Agricult.*, vol. 108, pp. 250–257, Oct. 2014, doi: 10.1016/j.compag.2014.08.011.

[117] D. L. Weller, T. M. T. Love, and M. Wiedmann, "Comparison of resampling algorithms to address class imbalance when developing machine learning models to predict foodborne pathogen presence in agricultural water," *Frontiers Environ. Sci.*, vol. 9, Jun. 2021, Art. no. 701288, doi: 10.3389/fenvs.2021.701288.

[118] Y. Cai, K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You, and B. Peng, "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches," *Agricult. Forest Meteorol.*, vol. 274, pp. 144–159, Aug. 2019, doi: 10.1016/j.agrformet.2019.03.010.

[119] J. A. Evans, P. J. Tranel, A. G. Hager, B. Schutte, C. Wu, L. A. Chatham, and A. S. Davis, "Managing the evolution of herbicide resistance," *Pest Manage. Sci.*, vol. 72, no. 1, pp. 74–80, Jan. 2016, doi: 10.1002/ps.4009.

[120] A. L. Diedrichs, F. Bromberg, D. Dujovne, T. Watteyne, "Prediction of frost events using Bayesian networks and random forest," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 2589–2597, Dec. 2018.

[121] M. N. M. Buthelezi, R. T. Lottering, S. T. Hlatshwayo, and K. Peerbhay, "Comparing rotation forests and extreme gradient boosting for monitoring drought damage on KwaZulu-natal commercial forests," *Geocarto Int.*, pp. 1–24, Dec. 2020, doi: 10.1080/10106049.2020.1852612.

[122] M. Khalaf, H. Alaskar, A. J. Hussain, T. Baker, Z. Maamar, R. Buyya, P. Liatsis, W. Khan, H. Tawfik, and D. Al-Jumeily, "IoT-enabled flood severity prediction via ensemble machine learning models," *IEEE Access*, vol. 8, pp. 70375–70386, 2020, doi: 10.1109/ACCESS.2020.2986090.

[123] M. Degiorgis, G. Gnecco, S. Gorni, G. Roth, M. Sanguineti, and A. C. Taramasso, "Classifiers for the detection of flood-prone areas using remote sensed elevation data," *J. Hydrol.*, vols. 470–471, pp. 302–315, Nov. 2012, doi: 10.1016/j.jhydrol.2012.09.006.

[124] G. Zhao, B. Pang, Z. Xu, and L. Xu, "A hybrid machine learning framework for real-time water level prediction in high sediment load reaches," *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124422, doi: 10.1016/j.jhydrol.2019.124422.

[125] W. Chen, X. Xie, J. Wang, B. Pradhan, H. Hong, D. T. Bui, Z. Duan, and J. Ma, "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility," *Catena*, vol. 151, pp. 147–160, Apr. 2017, doi: 10.1016/j.catena.2016.11.032.

[126] W. Chen, H. Li, E. Hou, S. Wang, G. Wang, M. Panahi, T. Li, T. Peng, C. Guo, C. Niu, L. Xiao, J. Wang, X. Xie, and B. B. Ahmad, "GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models," *Sci. Total Environ.*, vol. 634, pp. 853–867, Sep. 2018, doi: 10.1016/j.scitotenv.2018.04.055.

[127] O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran region, Iran," *Catena*, vol. 137, pp. 360–372, Feb. 2016, doi: 10.1016/j.catena.2015.10.010.

[128] B. T. Pham, A. Jaafari, I. Prakash, S. K. Singh, N. K. Quoc, and D. T. Bui, "Hybrid computational intelligence models for groundwater potential mapping," *Catena*, vol. 182, Nov. 2019, Art. no. 104101, doi: 10.1016/j.catena.2019.104101.

[129] S. Lee, Y. Hyun, S. Lee, and M.-J. Lee, "Groundwater potential mapping using remote sensing and GIS-based machine learning techniques," *Remote Sens.*, vol. 12, no. 7, p. 1200, Apr. 2020, doi: 10.3390/rs12071200.

[130] S. Miraki, S. H. Zanganeh, K. Chapi, V. P. Singh, A. Shirzadi, H. Shahabi, and B. T. Pham, "Mapping groundwater potential using a novel hybrid intelligence approach," *Water Resour. Manage.*, vol. 33, no. 1, pp. 281–302, Jan. 2019, doi: 10.1007/s11269-018-2102-6.

[131] P. K. Goel, S. O. Prasher, R. M. Patel, J. A. Landry, R. B. Bonnell, and A. A. Viau, "Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn," *Comput. Electron. Agricult.*, vol. 39, pp. 67–93, May 2003, doi: 10.1016/S0168-1699(03)00020-6.

[132] R. D. M. Fernandes, M. V. Cuevas, V. Hernandez-Santana, C. M. Rodriguez-Dominguez, C. M. Padilla-Díaz, and J. E. Fernández, "Classification models for automatic identification of daily states from leaf turgor related measurements in olive," *Comput. Electron. Agricult.*, vol. 142, pp. 181–189, Nov. 2017, doi: 10.1016/j.compag.2017.09.005.

[133] B. Barrett, I. Nitze, S. Green, and F. Cawkwell, "Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches," *Remote Sens. Environ.*, vol. 152, pp. 109–124, Sep. 2014, doi: 10.1016/j.rse.2014.05.018.

[134] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93–104, Jan. 2012, doi: 10.1016/j.isprsjprs.2011.11.002.

[135] P. A. Gutiérrez, F. López-Granados, J. M. Peña-Barragán, M. Jurado-Expósito, and C. Hervás-Martínez, "Logistic regression product-unit neural networks for mapping ridolfia segetum infestations in sunflower crop using multitemporal remote sensed data," *Comput. Electron. Agricult.*, vol. 64, no. 2, pp. 293–306, Dec. 2008, doi: 10.1016/j.compag.2008.06.001.

[136] M. J. Lambert and P. C. Traoré, X. Blaes, P. Baret, and P. Defourny, "Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt," *Remote Sens. Environ.*, vol. 216, pp. 647–657, Oct. 2018, doi: 10.1016/j.rse.2018.06.036.

[137] V. Lebourgeois and S. Dupuy, "É. Vintrou, M. Ameline, S. Butler, and A. Bégué, "A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM)," *Remote Sens.*, vol. 9, no. 3, p. 259, 2017, doi: 10.3390/rs9030259.

[138] H. Kerner, C. Nakalembe, and I. Becker-Reshef, "Field-level crop type classification with K nearest neighbors: A baseline for a new Kenya smallholder dataset," in *Proc. ICRL*, May 2020, pp. 1–6.

[139] D. Moshou, E. Vrindts, and B. De Ketelaere, J. De Baerdemaeker, and H. Ramon, "A neural network based plant classifier," *Comput. Electron. Agricult.*, vol. 31, no. 1, pp. 5–16, 2001, doi: 10.1016/S0168-1699(00)00170-8.

[140] M. B. Garzón, R. Blazek, M. Neteler, R. S. D. Dios, H. S. Ollero, and C. Furlanello, "Predicting habitat suitability with machine learning models: The potential area of pinus sylvestris L. In the Iberian peninsula," *Ecol. Model.*, vol. 197, nos. 3–4, pp. 383–393, Aug. 2006, doi: 10.1016/j.ecolmodel.2006.03.015.

[141] G. M. de Almeida, G. T. Pereira, A. S. R. de Souza Bahia, K. Fernandes, and J. M. Júnior, "Machine learning in the prediction of sugarcane production environments," *Comput. Electron. Agricult.*, vol. 190, Nov. 2021, Art. no. 106452, doi: 10.1016/j.compag.2021.106452.

[142] M. Ozdogan and G. Gutman, "A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental U.S.," *Remote Sens. Environ.*, vol. 112, no. 9, pp. 3520–3537, Sep. 2008, doi: 10.1016/j.rse.2008.04.010.

[143] R. G. Perea, E. Camacho Poyato, P. Montesinos, and J. A. R. Díaz, "Prediction of irrigation event occurrence at farm level using optimal decision trees," *Comput. Electron. Agricult.*, vol. 157, pp. 173–180, Feb. 2019, doi: 10.1016/j.compag.2018.12.043.

[144] S. Galelli and A. Castelletti, "Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling," *Hydrol. Earth Syst. Sci.*, vol. 17, no. 7, pp. 2669–2684, Jul. 2013, doi: 10.5194/hess-17-2669-2013.

[145] L. Ornella and E. Tapia, "Supervised machine learning and heterotic classification of maize (Zea mays L.) using molecular marker data," *Comput. Electron. Agricult.*, vol. 74, no. 2, pp. 250–257, Nov. 2010, doi: 10.1016/j.compag.2010.08.013.

[146] W. Okori and J. Obua, "Machine learning classification technique for famine prediction," in *Proc. World Congr. Eng.*, 2011, vol. 2, no. 1, pp. 991–996.

[147] Z. C. Lipton, "The mythos of model interpretability," *Commun. ACM*, vol. 61, no. 10, pp. 36–43, Sep. 2018, doi: 10.1145/3233231.

[148] B. Kim, R. Khanna, and O. Koyejo, "Examples are not enough, learn to criticize! Criticism for interpretability," in *Proc. Adv. Neural Inf. Process. Syst. (Nips)* 2016, pp. 2288–2296.

[149] C. Dau and D. Graff, "UCI machine learning repository," School Inf. Comput. Sci., Univ. California, Irvine, Irvine, CA, USA, 2019. [Online]. Available: https://archive.ics.uci.edu/ml

[150] A. Weller, "Transparency: Motivations and challenges," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (Lecture Notes in Computer Science), vol. 11700, W. Samek, G. Montavon, A. Vedaldi, L. Hansen, and K. Müller, Eds. Cham, Switzerland: Springer, 2019, pp. 23–40.

[151] M. Tulio Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," 2016, *arXiv:1606.05386*.

[152] O. Biran and C. Cotton, "Explanation and justification in machine learning: A survey," in *Proc. IJCAI Workshop Explainable AI (XAI)*, 2017, pp. 8–13. [Online]. Available: http://www.cs.columbia.edu/~orb/papers/xai_survey_paper_2017.pdf

[153] A. A. Freitas, "Comprehensible classification models," *ACM SIGKDD Explor. Newslett.*, vol. 15, no. 1, pp. 1–10, Mar. 2014, doi: 10.1145/2594473.2594475.

[154] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019, doi: 10.1073/pnas.1900654116.

[155] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, pp. 206–215, May 2019.

[156] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, "Network dissection: Quantifying interpretability of deep visual representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2018, pp. 6541–6549.

[157] L. Maimon and O. Rokach, "Decomposition methodology for knowledge discovery and data mining," in *Machine Perception and Artificial Intelligence*. Singapore: World Scientific, 2005.

[158] R. S. Michalski, "A theory and methodology of inductive learning," *Machine Learning*, R. S. Michalski and J. G. Carbonell, Ed. Berlin, Germany: Springer, 1983, doi: 10.1007/978-3-662-12405-5_4.

[159] C. Molnar, *Interpretable Machine Learning*. lulu.com, 2021. [Online]. Available: https://christophm.github.io/interpretable-ml-book/

[160] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?': Explaining the predictions of any classifier," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[161] G. Valdes, J. M. Luna, E. Eaton, C. B. Simone, L. H. Ungar, and T. D. Solberg, "MediBoost: A patient stratification tool for interpretable decision making in the era of precision medicine," *Sci. Rep.*, vol. 6, p. 37854, Nov. 2016, doi: 10.1038/srep37854.

[162] B. Kim, "Interactive and interpretable machine learning models for human machine collaboration," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Massachusetts Inst. Technol., Cambridge, MA, USA, 2015.

[163] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020.

[164] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1527–1535.

[165] D. Nauck, "Measuring interpretability in rule-based classification systems," in *Proc. 12th IEEE Int. Conf. Fuzzy Syst.*, May 2003, pp. 196–201, doi: 10.1109/FUZZ.2003.1209361.

[166] A. Cano, A. Zafra, and S. Ventura, "An interpretable classification rule mining algorithm," *Inf. Sci.*, vol. 240, pp. 1–20, Aug. 2013, doi: 10.1016/j.ins.2013.03.038.

[167] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability," *Soft Comput.*, vol. 13, no. 10, pp. 959–977, Aug. 2009, doi: 10.1007/s00500-008-0392-y.

[168] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011. [Online]. Available: https://scikit-learn.org/stable/about.html

[169] C. Meng, L. Trinh, N. Xu, and Y. Liu, "Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset," *Sci. Rep.*, vol. 12, p. 7166, 2021, doi: 10.1038/s41598-022-11012-2.

[170] A. A. Ismail, M. Gunady, H. C. Bravo, and S. Feizi, "Benchmarking deep learning interpretability in time series predictions," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2020, pp. 1–12.

[171] S. Hooker, D. Erhan, P. J. Kindermans, and B. Kim, "A benchmark for interpretability methods in deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019, pp. 1–12.

[172] T. N. Doan and T. A. Hoang, "Benchmarking neural topic models: An empirical study," in *Proc. Findings Assoc. Comput. Linguistics (ACL-IJCNLP)*, 2021, 4363–4368, doi: 10.18653/v1/2021.findings-acl.382.

[173] A. Y. Sun and B. R. Scanlon, "How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions," *Environ. Res. Lett.*, vol. 14, no. 7, Jul. 2019, Art. no. 073001, doi: 10.1088/1748-9326/ab1b7d.

[174] A. Khanna and S. Kaur, "Evolution of Internet of Things (IoT) and its significant impact in the field of precision agriculture," *Comput. Electron. Agricult.*, vol. 157, pp. 218–231, Feb. 2019, doi: 10.1016/j.compag.2018.12.039.

[175] J. Borchi, M. Woodcock, M. Redshaw, and B. Raniga, "Cyber security threats—Are we prepared? A threat based assessment of the cyber resilience of the Australian agricultural sector," AgriFutures Australia, Wagga Wagga, NSW, Australia, Tech. Rep. 21-070, Jun. 2021.

[176] D. Claughton and N. Beilharz. (Jun. 2021). *JBS Foods Pays $14.2 Million Ransom to End Cyber Attack on Its Global Operations*. [Online]. Available: https://www.abc.net.au/news/rural/2021-06-10/jbs-foods-pays-14million-ransom-cyber-attack/100204240

[177] E. Bout, V. Loscri, and A. Gallais, "How machine learning changes the nature of cyberattacks on IoT networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 248–279, 1st Quart., 2022, doi: 10.1109/COMST.2021.3127267.

[178] I. Rosenberg, A. Shabtai, Y. Elovici, and L. Rokach, "Adversarial machine learning attacks and defense methods in the cyber security domain," *ACM Comput. Surv.*, vol. 54, no. 5, pp. 1–36, Jun. 2022, doi: 10.1145/3453158.

[179] R. Hu, Z. Yan, W. Ding, and L. T. Yang, "A survey on data provenance in IoT," *World Wide Web*, vol. 23, no. 2, pp. 1441–1463, Mar. 2020, doi: 10.1007/s11280-019-00746-1.

[180] Y. Al-Hadhrami and F. K. Hussain, "DDoS attacks in IoT networks: A comprehensive systematic literature review," *World Wide Web*, vol. 24, no. 3, pp. 971–1001, May 2021, doi: 10.1007/s11280-020-00855-2.

[181] T. McIntosh, A. S. M. Kayes, Y.-P.-P. Chen, A. Ng, and P. Watters, "Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–36, Dec. 2022, doi: 10.1145/3479393.

[182] M. Bewong, J. Liu, L. Liu, and J. Li, "Privacy preserving serial publication of transactional data," *Inf. Syst.*, vol. 82, pp. 53–70, May 2019.

SARAH CONDRAN is currently pursuing the Ph.D. degree in computer science with Charles Sturt University. She has previously worked on data analysis to benchmark crop performance within the potato sector in the U.K. Her current research interests include applied machine learning, text analytics, and information warfare.

MICHAEL BEWONG (Member, IEEE) received the Ph.D. degree from the University of South Australia, Adelaide, Australia. He is currently a Lecturer in computing at Charles Sturt University, Australia. He has worked on various industry-based applied machine learning projects. He has also authored top tier papers in this area. His current research interests include data mining, machine learning, and cyber security.

MD ZAHIDUL ISLAM is currently a Professor of computer science with the School of Computing, Mathematics and Engineering, Charles Sturt University, Australia. His main research interests include data mining, classification, and clustering algorithms, missing value imputation, data cleaning and pre-processing, privacy preserving data mining, privacy issues due to data mining on social network users, and applications of data mining in real life.

LANCELOT MAPHOSA received the Ph.D. degree. He is currently a Research Scientist at the New South Wales Department of Primary Industries, Wagga Wagga, Australia. His main research interests include modern genotyping and phenotyping approaches, data analysis, quantitative genetics, agronomy, and crop physiology.

LIHONG ZHENG (Member, IEEE) received the Ph.D. degree in computing sciences from the University of Technology Sydney, in 2008. She is currently an Associate Professor at the School of Computing, Mathematics and Engineering, Charles Sturt University. Her previous research interest was on automation, robotics, and artificial intelligence. Her current research interests include computer vision, machine learning, image processing, pattern recognition, wireless sensor networks, and network security.