# A Survey of Machine Learning Modelling for Agricultural Soil Properties Analysis and Fertility Status Predictions

Augustine Josephat Josephat Malamsha [*] , Mussa Ally Dida , Sabine Moebs

*Review*

# A Survey of Machine Learning Modelling for Agricultural Soil Properties Analysis and Fertility Status Predictions

**Augustine Josephat Malamsha [1],*, Mussa Ally Dida [1] and Sabine Moebs [2]**

[1] School of Computational and Communication Science and Engineering, Nelson Mandela African Institute of Science and Technology, P.O.BOX 447 Arusha, Tanzania; malamshaa@nmaist.ac.tz (A.J.M); mussa.ally@nm-aist.ac.tz (M.A.D)

[2] Department of Digital Business Management, Baden-Wuerttemberg Cooperative State University Heidenheim, Marienstraße 20, 89518 Heidenheim an der Brenz, Germany; sabine.moebs@dhbw-heidenheim.de (S.M)

**\*** Correspondence: malamshaa@nmaist.ac.tz; Augustine.malamsha@ifm.ac.tz; Tel.:+255766280341

**Abstract:** The problem of low soil fertility and limited research in agricultural data driven tools, may lead to low crop productivity which makes it imperative to research in applications of high throughput computational algorithms such as of machine learning (ML) for effective soil analysis and fertility status prediction in order to assist in optimal soil fertility management decision-making activities. However, difficulties in the choice of the key soil properties parameters for use in reliable soil nutrients analysis and fertility prediction. Also, individual ML algorithms setbacks and modelling expert implementation procedures subjectivity, may lead to exploitation of worst fertility level targets and soil fertility status targets classification models performance reported variations. This paper surveys state-of-affair in ML for agricultural soil nutrients analysis and fertility status prediction. Prominent soil properties and widely used classical modelling algorithms and procedures are identified. Empirically exploited fertility status target classes are scrutinized, and reported soil fertility prediction model performances are depicted. The three pass method, with mixed method of qualitative content analysis and qualitative simple descriptive statistics were used in this survey. Observably, the frequently used soil nutrients and chemical properties were organic carbon, phosphorus, potassium, and potential Hydrogen, followed by iron, manganese, copper and zinc. Predominant algorithms included Random Forest, and Naïve Bayes, followed by Support Vector Machine. Model performances varied, with highest accuracy 98.93% and 98.15% achieved by ensemble methods, and the least being 60%. Interdisciplinary ML related researchers may consider using ensemble methods to develop high performance soil fertility status prediction models.

**Keywords:** artificial intelligence; machine learning; soil nutrients analysis; soil fertility prediction; smart soil fertility management; smart farming

## 1. Introduction

Machine Learning (ML) is a field of study that shows great promises in the interdisciplinary study of agricultural soils analysis through modelling for knowledge discovery such as in soil fertility predictions task(s) with performances that approach expertise of soil laboratory technicians, but at lowered costs-and wider magnitude of application. ML modelling for soil fertility predictions is vital since the soil is the key fundamental factor for crop growth and productivity, other than crop diseases, irrigation, weather and climatic conditions management, as it harbors plantations and as a function of its fertility it can supply plantations with the necessary nutrients as well as other chemical and biological possessions necessary for plant growth and increased crops yields, to eventually ensure food security [1,2]. Shown in Figure 1 is a diagram synthesizing from [3], that depicts of the study rationale adapted a synthesis for depicting soil and key components with properties necessary for plant growth to be proposed for integrated with ML for provision of smart agricultural system.
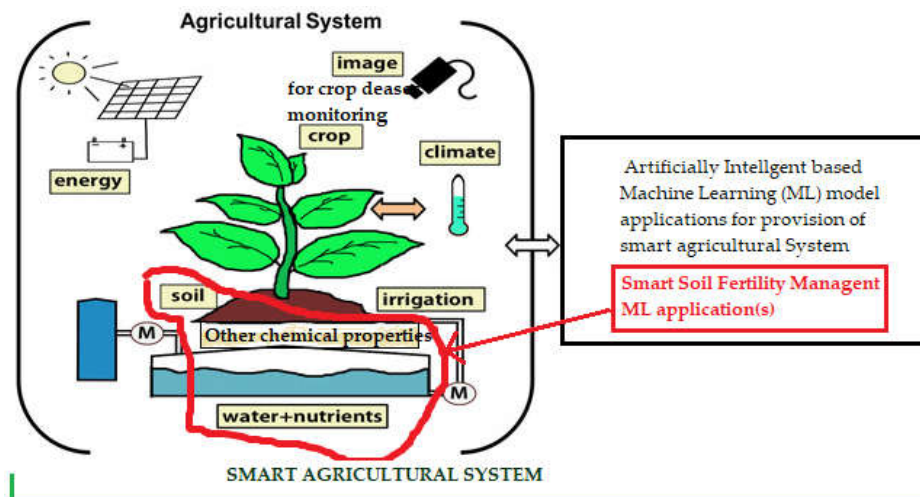
**Figure 1.** A depiction of soil and key components with properties necessary for plant growth to be proposed for integrated with ML for provision of smart agricultural system.

As a matter of fact, due to both natural and human causes such as: wind and soil erosion, unfavorable climatic conditions, extreme weather conditions, mono-cropping or monoculture, total clearance of crop residues, intensive tillage, draught, non-fine-tuned (or non-site specific  fertility conditions considered) use of fertilizers, the soil may become highly susceptible to the problem of low soil fertility characteristic. In turn, this  may highly results into low crop yields which is clearly a threat to food security  [2,4,5]. Thus, if limited efforts in research and developments that are related to ML solutions deployments for smart agricultural soil fertility management to deal with the low soil fertility characteristic problem, such as the one depicted in Figure 1 will be the fact to prevail, then it may become more difficult to have provisions for fine-tuned or site specific soil fertility restoration or preservation measures of fertilization, crop rotation, non-tillage farming, mixed planting, sowing green manure, mulching, and fallowing, this of which is necessary to increase field fertility [5]. Most critically, the infertility problem even cause catastrophic hunger condition(s) by the year 2050, when the world's population growth will reach the estimated 9.6 to 9.73 billion people(approximately 10 billion) from the current 7.3 billion [6–9]. Therefore, to achieve smart soil fertility management system as part of a sustainable smart global food production and supply system which is currently a major demand of the United Nations Food and Agriculture organization (FAO) is apparently imperative. Thus, research and development of high throughput ML related models applications for agricultural soil analysis to assist in effective soil fertility management decision-making processes becomes obviously imperative [2,4], especially in the context of highly susceptible regions.

Nevertheless, the choice of deployment or developments of subtle ML models for soil fertility nutrients and other chemical properties is quiet challenging as there exists a multitude of agricultural soil properties that can be used as features for modelling soil fertility status predictors. This may be because of subjectivity arising from the involvement and bases of associated experts context that provide necessary prior knowledge for the deemed fertility levels targets for the data required to implement these models. Also, a choice of appropriate ML algorithms and implementation approaches and procedures that may fit best models in the available data context is one of key challenges, whereby these factors may lead to impaired overall model performances, consequently hindering these models practical or operational application in real world decision making processes.

Therefore, this paper aims to survey the most current state of ML studies related to research in ML models for soil fertility nutrients and other chemical properties modelling for predicting soil fertility status. The paper first provide from an empirical perspective, the key soil nutrients and its other chemical properties these of which are directly significant to crop growth and depiction of soil fertility[87], as key features for a ML interdisciplinary study of soil fertility status prediction. Secondly, it also identifies major ML algorithms for use in modelling these soil features. Then, it

depicts the soil fertility target classes these of which are crucial in providing for fine-tuned treatments of soil shall the need arise. Finally, the survey summarizes soil fertility prediction models performances, this which are necessary to provide farmers and other relevant agricultural stakeholders with models that provide optimal analytical information and recommendations for use in decision making about sites specific fertilizer dosages[10–12]. This survey is generally useful to future researchers and soil assessment entities as amongst others things it helps in the choice of key soil properties for soil fertility assessment, this of which have been a primary concern since the beginning of soil management assessment frameworks (SMAF) and analysis tools developments programmes[4].

The rest of this paper is organized as follows: Subsection 1.1 describes soil properties. Subsection 1.2, selectively briefs on some data source that are related to the Soil properties described in subsection 1.1. Subsection 1.3 overviews the machine learning discipline, with a briefing on its operational techniques and algorithms, as well as its classification models major performance evaluation metrics. Section 2 provides impacts of data-driven ICT with respect to use of machine learning in agriculture. Section 3 is about materials and methods used to conduct this survey. Section 4 provides the empirical results about machine learning for soil nutrients analysis and fertility status prediction as observed from previous studies. Section 5 is a summary of the findings attained by this survey and discussion of those findings. Finally, section 6 concludes the survey and provides recommendation for future studies.

### 1.1. Soil Properties

With respect to the agricultural soils, there are key properties which are functional to estimate its fertility, propose a cultivation plan and predict crop productivity include its soil physical, chemical and biological properties [4,11]. These should be selected based on some criteria such as the soil properties selection criteria of conceptuality or operation ability, practicability or availability, sensitivity, interpretability [4,13], and they can then be determined through various field trials (treatments) of soil sampling from different depth, mostly sub and upper soil, then testing of the samples by using several different soil testing methods to determine availability and extent of different soil properties [4,13].

Soil physical properties which amongst others includes its structure: texture or type as clay, loamy, sandy, silty, peaty, chalky or mixture of these, water storage capacity, porosity, and infiltration [4]. Most importantly, according to [4,11,14], soil fertility can be determined by chemical properties with which are mostly nutrients that can be determined by using either wet or dry chemistry through techniques such as Unnamed Arial Vehicles (UAVs), near or mid infrared (NIR/MIR) spectroscopic method and calibrated using samples that underwent a wet chemistry method with reagents such as HCLO4, HCL, HF and HNO3 [15], sodium hydrogen carbonate extraction, fulvic acids (FA), Humic acids (HA), BaCl2 extraction, Dionex-100 Ionic Chromatography (DX 1-03, USA) [13,16], amongst others, as suggested by [4], these soil chemical properties include total nitrogen content (N) in the form of nitrous oxide – N20, phosphorus (P) in the form of phosphorus pentoxide – P205, potassium (K) in the form of potassium oxide – P20, sulphur (S) in sulphate – SO4 form, iron (Fe), manganese (Mn), copper (Cu), zinc (Zn), boron (B), magnesium (Mg), calcium in calcium carbonate – CaCO3, cation exchange; organic carbon (OC) measured in percentage (%); electrical conductivity (EC) measured in milliSiemens per meter (mS/m) or deciSiemens per meter; potential hydrogen (pH) measured by a pH meter as acidic (0 to less than 5.5), neutral (5.5 to 7) or alkaline (7 to 14). Also other core soil chemical properties are the amount salt dissolved in water or saltiness, the salinity measured in ppm or %; and the amount of water contained in the soil, moisture, which is measured in percentage (%) of wet to dry soil [4,11,14]. Biological properties provide the key living component of the soil ecosystem, microorganisms that decompose organic matter into top soil organic carbon responsible for the preservation of soil nutrients and formation of soil natural fertility measured in percentage (%) of total Nitrogen [4]. Additionally, there are some other agronomic parameters which are not part of the three soil properties categories but can be

essential indexes for the estimation of soil fertility; they include crop type as seen in the work of [4], and yield amounts [4,11,14].

Therefore, soil being the fundamental key agricultural aspect that harbors nutrients for crop consumption and growth [1,17–19], in this advent of big data, including those of agricultural soils, it is imperative to research in and apply high throughput analytical techniques [20], such as those of machine learning in soil nutrients analysis and fertility status prediction problems, in order to intelligently assess and management the soil through accurate and precise determination of farm fields site specific nutrients and other temporal and spatial properties variabilities, and support for decision advices on the right site specific fertilizer doses, treatments, and required soil management practices [14].

### 1.2. Soil Information Data Source

Studies have widely reported that, accurate analysis and predictions in agriculture can be achieved through the application of data-driven tools such as those using machine learning algorithms as they have the ability to unlock and reveal patterns of hidden knowledge. [2,21] asserted that, the use of machine learning algorithms in this advent of large collected data for example can manipulate data and produce knowledge required for making better precision agriculture and support decision-making among farmers and other agricultural stakeholders, such as in soil assessment and management.

However, despite that a lot of soil data is being collected worldwide and in Africa through various projects like the Sub Saharan's Africa Soil Information Services (AfSIS) and Ethiopia's, Nigeria's and Tanzania's Taking Maize Agronomy to Scale in Africa (TAMASA), and, the Tanzania's specific Tanzania Soil Information Services (TanSIS) [22]; with which this data can be used as input to machine learning in the development of models such as those for the national soil quality assessment and soil management assessment frameworks (SMAF) [4] or tools in support of advanced soil analysis tasks and nutrients mapping [2].

### 1.3. Machine Learning

Machine learning (ML) is a branch of artificial intelligence (AI) which has brought about many advancements in application areas such as robotics, natural-language processing, expert systems, and ML, it is a subfield of computer science that focuses on the design and development of intelligent systems in the form of hardware, software, or both [23,24].

Whereas Machine engineering and learning is concerned with the design and development, and application of algorithms and techniques that learn by automatically organizing input data according to their common features to provide computer machine(s) with knowledge and experience in the form of mathematical models that infer the future with minimal human intervention thus with least errors and make decisions [25–29], whereby ML does not involve consciousness as by humans, rather it statistical regularities or other data patterns, thus ML hardly resembles human approaches to learning [30]. Consequently ML becomes one of the key aspects of an AI agent it provide for knowledge discovery computational structures or model of an AI system, pattern recognition, and data mining by   learning from data to discover computational structures these of which can later on be used for predictive and descriptive work flows termed as improved performance analytics [31]. Thus, the key take away of ML is when performance of the machine improves with even a slight change of any aspect of an AI system then the machine is said to have learned.

Broadly speaking, a machine is said to learn when it changes its structure, program or data, such that its performance in that carrying out artificial intelligence related tasks that involves diagnosis, prediction, robotic control, planning, recognition, improves with changes for enhancing its system or ab initio synthesis of new systems, as portrayed by [32]. With respect to machine learning as a data science discipline, its wide applications have been observed in different fields ranging from transportation, medical health, education and agriculture, and have been reported to tremendously outperform humans as   well as conventional computer programs performances by propelling very solid results in all tasks, for example they could achieve 98.98% accuracy, higher than humans at

traffic signs [33]. In medical health and education, machine learning have been relatively used for the detection of heart and breast cancer [34]; drug design [35], and rational drug discovery [36]; predicting student performance [37,38], and dropouts [39,40].

### 1.3.1. Machine Learning Modelling Techniques and Algorithms

In machine learning there are dozens of machine learning algorithms that fall into either of four main classical ML modelling techniques known as supervised, unsupervised, reinforcement (semi-supervised) [41,42]. Furthermore, there came an emergence of Deep ML which have been reported to have the capability of achieving the most highest performances than any other learning technique, including classical machine learning techniques and its variants [25,43–45]. However, deep learning techniques are associated with drawbacks and limitations which are respectively high processing times and need for very large image data [25,43,45], while existing soil testing methods provide textual dataset contrary to images, therefore deep learning cannot currently be applied for the analysis and prediction of soil nutrients and fertility status, otherwise we need to set up project(s) for the collection of very large soil properties data in the form of images, with which this involves high cost and time. In the context of agriculture on addressing soil nutrients analysis and fertility status prediction, based on the nature of agricultural soil data which is mostly textual, the techniques for learning can be supervised or unsupervised, with supervised being widely used for soil nutrients analysis and fertility prediction as observed from the literature.

Supervised machine learning is characterized by teacher or supervisor with the task to provide an agent, model or function with a precise measure of its own errors, whereby beliefs and common sense presented in the form of training data set made up of inputs and expected outputs or class labels are provided, and the function shall be used to infer for future unseen samples, thus the function will map a vector into a specific class from the several by looking at the functions input-output sets of examples [25,30]. Through the application of various ML algorithms, supervised learning has been reported to be efficient in finding solutions to several linear and non-linear problems such as predictive analysis based on regression or categorical classification. Generally, the traditional supervised learning algorithms uses a training set D with variables constituting of predictors X and target Y to train a model. The training process seeks to identify through an iterative procedure a set of model parameters that maximizes the relationship between the predictor (input features) and the target variables. The trained model receives new input data for the predictor variables and uses the recognized pattern to estimate the target variable. If the datasets used contains continuous values and created function produces   continuous valued outcomes, the task is refers to as regression, otherwise if they are based on discrete number of possible outcomes, then it is a classification problem [25,30]. Also, when the algorithm used to create the function is flexible enough and data is coherent, the overall accuracy increase and predicted to expected values difference closes nearly to zero, the goal is to reduce the number of misclassifications and increase robustness to noise.

Unlike supervised, Unsupervised learning algorithms are used to identify hidden patterns in unlabeled input data; they refers to provide ability to learn and organize information without an error signal and be able to evaluate the potential solution, this type of learning simply models a set of inputs with no labeled examples [25,30]. The lack of direction for the learning algorithm in unsupervised learning can sometime be advantageous, since it lets the algorithm to look back for patterns that have not been previously considered [25,46]. In unsupervised learning, training is conducted using dataset *D* without function and we aim to partition the training set *D* into subsets, *D1, . . . , DR*, in an appropriate manner. Whereby the value of the function is the name of the subset to which an input vector belongs. In some cases when unsupervised results are to be used as inputs into a supervised process, problem domain expert(s) intervention becomes valuable for additional verification of the unsupervised learning intermediate results for enhanced performance and reliability, such as in the verification of different an unsupervised environment created clusters or otherwise termed as class labels to be used in supervised learning [47].In some cases when unsupervised results are to be used as inputs into a supervised process, problem domain expert(s) intervention becomes valuable for additional verification of the unsupervised learning intermediate

results for enhanced performance and reliability, such as in the verification of different unsupervised environment created clusters or otherwise termed as class labels to be used in supervised learning [47].

Literally, both supervised and unsupervised learning are associated with a wide range of algorithms for use in modelling either of the related respective ML tasks. [48] asserted that, thousands of machine learning algorithms exist and hundreds more are being published each year. Long before the 90's, theories of conventional statistics have been existing along with core machine learning artificial neural networks (ANNs) and decision trees (DTs) based techniques, these of which have been widely applied in field of medical research for effective drug discovery problem tasks. In a broader view, one remarkable discovery in machine learning algorithms was in 1986, when the induction decision tree (ID3), a variant of the native DT algorithm which models data in tree like structures as its name suggests[29,30,49]. ID3 was proposed by Quinlan as an algorithm with the potential to provide transparent and interpretable explanations in the underlying rules of the tree structure, clearly stating reasoning behind reaching certain conclusions [33,36,50]. That of which is one of the key advantages of these algorithms, that is their simplicity and comprehensibility to determine small or large data structures, or attributes that provide information to solve classification and regression predictive problems. Quinlan later on further developed an improvement of the ID3 to form the C4.5 whose Java version is known as J48 [51–53].

One of the biggest booms that sparked machine learning was based on the neuropsychological learning formulation that mimic human brain functioning to create of artificial neural network(ANN) algorithm. ANN is an attractive and powerful algorithm initially highly used to model drug discovery researches as at 1995. The algorithms operates by determining  and minimizing errors through the network adjustments 30,54–56] by using network structures that can mainly be classified into four main topological approaches, namely, feed forward neural networks (FFNNs), backward propagation neural networks (BPNNs), random neural networks (RNNs) and self-organizing neural networks (SONNs) [36]. As asserted by [57], BPNNs is the most popular ANN topology, and the forward neural network contains multilayered perception and uses the gradient-descendent method to minimize the mean-square errors of the difference between the experimental or training data set and the network outputs. The key advantage of ANN algorithms is its ability to detect all possible interactions between predictors variables without having doubts even in cases of complex nonlinear relationship between independent and dependent variables, this of which is one of their key advantages 30,54–56]. While, its drawbacks that arises from the forward and back propagations requirements of large computational times, especially if a lot of middle hidden layers are involved in the learning process leading to the vanishing gradient otherwise termed as gradient loss or descent problem. This of which leads to redundant learning hops after a certain period of time in such they are inclined to over-fitting in shot number of hops [58,59]. Also another drawback of ANN is requirement of very large amounts of training dataset, and its black-box nature, that is, the inability to provide explanation of the underlying facts for reaching conclusions[33,55]. This of which sparked the need for more explorations by the machine learning research and development community. As a result some remarkable machine learning algorithms were discovered, including the support vector machine (SVM), and later on the random forest (RF) to encounter, among others, the mentioned problems of gradient loss, over-fitting, outlier susceptibility, and black-box characteristics.

Another great breakthrough in machine learning was the introduction of SVM in 1995 [33,36,60,61], with its kernel version being released near the year 2000 making competition with the ANN community a bit more subtle. The SVM could exploit knowledge of convex optimization, generalization margin theory and kernels, with stronger theoretical standings and empirical results [33,60]. SVM are capable of handling small data sets having high-dimensional variables. These algorithms maps points in space to create separates categories by maximizing the margin between different classes of points in linear problems [36,62], while they use kernel mapping to transform nonlinear data sets into a high-dimension feature space that can be used in linear classification functions [36].

Subsequently in the year 2001, came the discovery of the random decision trees forest which is commonly known as random forest (RF) algorithm that. The RF had improved performances and robustness towards over-fitting and outliers as compared to not only the individual decisions trees, but also the adaptive boosting algorithms [33,36,63–65]. These algorithms works   by using a random selection of features and 'bagging idea', that is, constructing an ensemble of multiple decisions trees as base learners and training them using random sampled subsets of original dataset, a consensus score is calculate as a weighted average or estimate of the individual decisions trees output to provide the final result [33,36,63–65]. As adapted from 66], Figure 2 shows a non-exhaustive (Non-Exh.) taxonomy of some classical ML algorithms, which also depicts the non-prescribed herein clustering techniques with the respective highlighted K-means, density, hierarchical based algorithm, as well as the dimensionality which includes PCA, UMAP, and   tSNE algorithms of the unsupervised ML techniques.
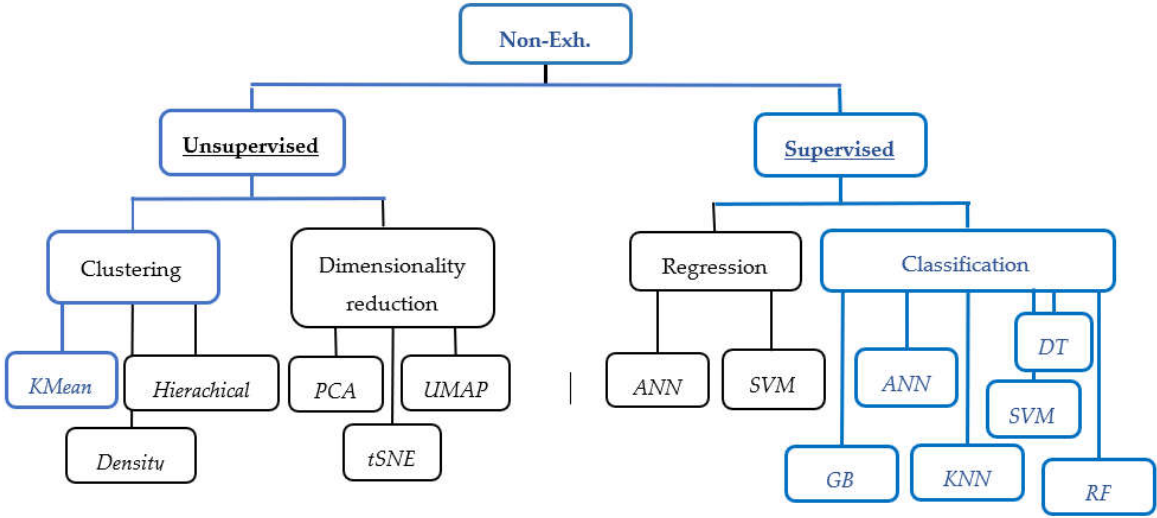


**Figure 2.** Non-Exhaustive Taxonomy of some classical ML algorithms.

Whereby PCA, tSNE, and UMAP are respective short forms for principle component analysis, t-distributed stochastic neighbor embedding, and uniform manifold approximation and projection.

However, following the era of prescribed classical machine learning techniques encountered the deep machine learning approaches, whereby classical ANN community decided not to remain behind its SVM, and RF rivals, and utilized the advents of big data and computational powers with to generate ANNs topological structures in hierarchical representation that allows larger learning capabilities to obtain highest performances and precision through deep machine learning algorithms such as deep neural networks (DNN), recurrent neural networks (RNN), deep belief networks (DBN) and convolutional neural networks (CNN) [36,43,44,67,68].

Now, based on machine learning algorithms performances, corresponding data-driven tools can be developed to effectively and efficiently unlock the potential for accurate and precise predictions and analysis of soil through estimation of key soil parameters and provide for decision support [1,2,4], in such to provide high agricultural yields and release laboratory soil scientists and farmers off their agricultural activities, so to say scientists and farmers can eventually obtain more time to concentrate on other agricultural economic activities like diversification of agricultural products [14].

### 1.3.2. Performance Evaluation

Measuring of the discrimination ability of a model is one of the   important aspect in assessing its performance [69]. This remains to be true following the "no free lunch theorem", whereby one cannot fit all. Therefore, the ML algorithms described in subsection 1.3.1 cannot apply and perform equally with similar performances in all problem domains areas and tasks, due to the fact that they will be subjected to different data complexities, let alone the individual algorithms drawbacks such

as need for large dataset or under lacking in their operational structures. Therefore, an evaluation of each algorithm model should be conducted to measure its performance as observed using metrics values. Based on the particular ML modelling task, either predictive supervised learning or descriptive unsupervised learning the appropriate performance metric will be selected. Thus because the main focus herein was on prediction of soil fertility status which is clearly a classification problem. With that respect, in this survey some major supervised learning algorithms models metrics are mostly described. These includes but are not limited to: the classifications accuracy, precision, recall, F1-measure also known as F1-score. Also receiver operating characteristics (ROC) analysis area under the curve (AUC) {ROC-AUC}, and Cohen kappa [70–72,72,73]. These of which are computed by using entries of a confusion matrix which mainly provides the number of true positive (TP), true negatives (TN), false positive (FP), and false negatives (FN) classifications by the ML classifiers models as shown in Table 1.

**Table 1.** The four basic confusion matrix metrics.

| Classes | Test result positive (+ve) | Test result negative(–ve) |
|---|---|---|
| Actual +ve | True +ve(TP) | False -ve (FN) |
| Actual –ve | False +ve (FP) | True -ve (TN) |

Classification Accuracy, is the percentage of correct predictions where the top class (the one having the highest probability), as indicated by the model, is the same as the target label. For multi-class classification problems, accuracy is averaged among all the classes. Accuracy is mentioned as Rank-1 identification rate [74]. This metric, accuracy, is the most intuitive performance metric for ML classification problem [75,76]. Mathematically, accuracy represents the ratio between some of the true positives to sum of true positives and false positive, that means the total observation, see equation (1)

$$Accuracy = (TP+TN)/(TP+FN + FP+TN), \qquad (1)$$

However it is mostly used to only provide the initial baseline performances and cannot be highly relied on. In such as derived from the confusion matrix more comprehensive metrics that are precision, recall, and f-measure are the mostly for evaluating the performance of machine learning classifiers [77].

Precision is the fraction of TP from the total amount of relevant results, that is, the sum of TP and FP, and it is mathematically presented by equation (2). For multi-class classification problems, precision is averaged among the classes[77].

$$Precision=TP/(TP+FP), \qquad (2)$$

Next is recall, defined as faction of TP from the total amount of TP and FN, and mathematically presented by equation (3). For multi-class classification problems, R gets averaged among all the classes[77].

$$Recall=TP/(TP+FN), \qquad (3)$$

Again, there is F1 Score or F-Measure, these are the harmonic means of precision and recall, which is mathematically presented by equation (4). For multi-class classification problems, F1 gets averaged among all the classes. It is mentioned as F-measure [77,78].

$$F1 = 2 * (TP * FP) / (TP+FP), \qquad (4)$$

Whereas, the ROC-AUC and Cohen kappa winds up the basic major metrics for classification problems which is the main focus of the problem task in this survey, that is to predict soil fertility status. ROC-AUC is widely used to determine these models discrimination ability [71], and Cohen kappa applies in measuring the closeness at which machine learning classified instances matches ground truth data labeled, also termed as agreement between two raters [79,80]. For, regression problems, Root Mean Square Error (RMSE) and Mean Square Error (MSE) which is the errors between

predicted and observed continuous outcomes from a regression model, are the widely used metrics, amongst others.

Contrary to supervised learning, unsupervised learning such as clustering problems uses metrics such as distance measure, the within group similarities and distances between segments or clusters or otherwise termed as groups [47]. And, the convergence or time taken to by the algorithm to bring out a model remains a vital metric for any machine learning related task[43,75,76]. Although past performance may not be indicative of future results, the mentioned metrics forms the common ground for determining how well the developed models might perform in the future. These metrics can mostly all be computed by using the powerful Scikit-learn (Sklearn) libray for ML applications in python [81], that consists of a wide range of ML classification metrics, and other machine learning modelling task other supervised learning, like the regression metrics highlighted herein this survey paper.

## 2. Data-Driven ICT and Impact of Machine learning in Agriculture

One of the earliest quoted example of the applications of machine learning in agriculture was in the use of similarity based learning to identify rules for the diagnosis of soy bean disease [82]. Other studies of the machine learning field in agricultural problems includes weather forecasting, yield prediction [83–85], fertilizers usage , fruit grading, plant diseases diagnosis and prediction [82], pest management [86], weed detection , soil nutrients analysis and fertility prediction for soil management and assessment [17,87], amongst others. Studies related to soil analysis and fertility status prediction have widely reported that accurate soil fertility estimation and predictions models in agriculture can be achieved through the application of data-driven tools such as those using machine learning algorithms, as these have the ability to unlock the such potentials.[2,21] asserted that, the use of machine learning algorithms in this advent of large collected data for example can manipulate data and produce knowledge required for making better precision agriculture and support decision making among farmers and other agricultural stakeholders, such as in soil assessment and management. Whereas much agricultural ML research and application have been conducted by developed regions, developing countries have limited research in the similar[2]. Consequently that hampers agricultural populations in these agricultural ML research under lacking regions from gaining better understanding of their farm fields conditions, such as determination of site-specific nutrients deficiencies for appropriate remedies to take place including application of the right fertilizer dosage [19,88–90].

In addition to the developed countries being pioneers of early developments and improvements of the profound agricultural and soil assessment trends, methods, tools and frameworks. For many years they have been researching in the development and applications of ICT data-driven tools such as of machine learning to implement and use precision or smart agriculture in various agricultural activities. In turn, this is a key supplement for automated business intelligence system for management decision making processes such as those that are geared towards improvement of soil fertility through soil nutrients advanced ML modelling and analysis, to realize automation in farming activities leading to increased yields while preserving the environment for sustainable agricultural through the optimal use of agricultural resources [91].

By using the agricultural gross domestic products (AGDP) and labor force (AGLF) by population data that is reported by [92,93], the respective ratios could be calculated as shown in Figure 3 of the Agricultural gross domestic products (AGDP) to labor force (AGLF) ratios. It can be deduced that limited machine learning (ML) related research and development could highly be among the reasons that leads to smaller ratios of agricultural gross domestic product to agricultural labor force in developing countries as compared to developed countries.
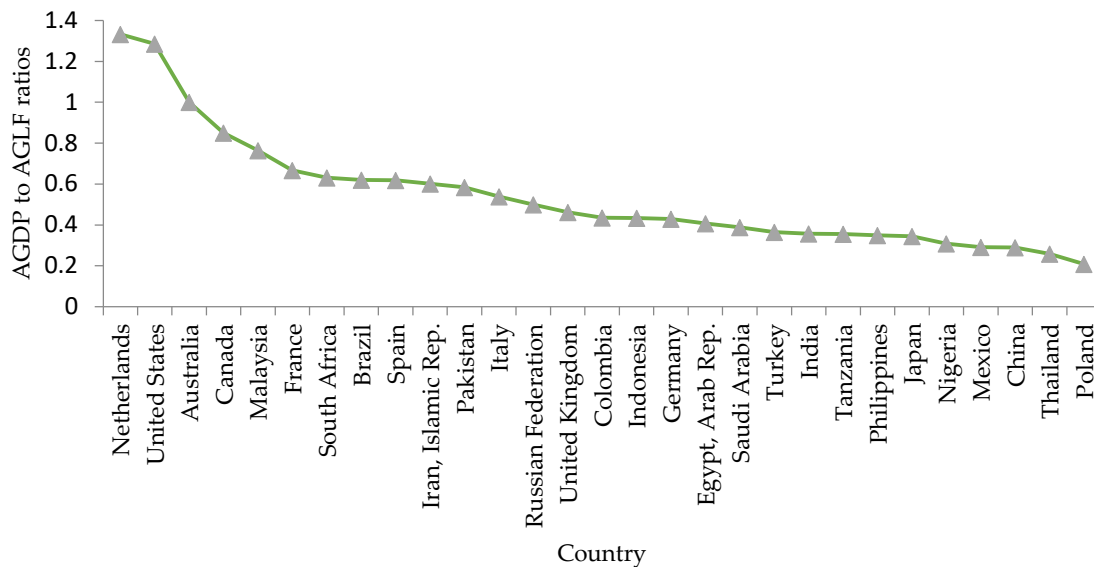
**Figure 3.** Ratios of ALFbP to AGDP.

Among other reason leading to the lowered AGDP to AGLF ratios in the developing countries being due to limited research in ML related application for smart agricultural systems implementations. Thus, for most of regions with low digital agriculture adoption, it may become imperative to research in development and application of machine learning (ML) techniques and models [20], for the development of agricultural data-driven tools for effective soil analysis. As part of a smart fertility management practices, this would be necessary for relevant applications such as for soil fertility status prediction, these of which can intelligently analyze soil nutrients and other chemical properties which have a direct interlink to crop productivity. In such, site specific farm field's accurate predictions of soil fertility status can be obtained to provide a reliable understanding of nutrients and other temporal and spatial soil variabilities or deficiencies that would require appropriate remedies. This of which can mainly be possible through the use of agricultural soil data-driven assisted approach for assisting in optimal soil related decision making processes during preparations phase before plantation, through ML model based recommendation or advices for appropriate site specific treatments including the required fertilizer dosages, and management practices [14].

## 3. Materials and Methods

This work followed a semi-systematic approach and was accomplished by the help of a paper review matrix with an extensive three pass method after collecting, organizing, and reviewing of pertinent materials as suggested by [94,95]. Moreover, the six W6H framework's interrogative questions of what, why, when, who, where and How, were used to effectively yield extra and necessary information to assist in the engineering of concepts and creation of a holistic architectural enterprise of machine learning, and application in agricultural soil analysis, as part of this survey [96–98].

Except for the four articles i) "The organization of behavior New York" by [99] which highlighted on the initial emergence of core Hebbedian learning theory,  ii) "Learning representations by back-propagating errors" by [57] which enlighten about the popular back propagation neural network (BPNN) machine learning method in existence long before the 1990's, iii) "Induction of decision trees" by [33,36,50] which proposed a remarkable discovery of the ID3 algorithm as a transparent and interpretable algorithms that can explain the underlying rules which are black-boxed in former algorithms and theories, and iv) "Learning by being told and learning from examples: an experimental comparison of the two methods of knowledge acquisition in the context

of development an expert system for soybean disease diagnosis" by [82] which highlighted on the earliest quotation in the application of machine learning in agriculture; The total formulation of this survey was accomplished through a review of high scoring pertinent online materials that were mostly published during the period of the past three decades starting from inclusively 1990 to 2019. The choice of publications starting from the year 1990 is based on the fact that during the early 90's, the classical machine learning family gained popularity, made a lot of achievements and discoveries, were commercialized on personal computer and shifted work from knowledge to data driven approach, and scientists began creating computer programs that analyzed large amounts of data and draw conclusions (learn from results) [65,100]. In addition, for the main focus area of this survey, that is respects to machine learning field in the analysis and fertility prediction of agricultural soils, in order to have the most recent state of affair, the survey utilized research works conducted from the year 2000, with the majority of this work starting inclusively from 2010, with 2015 marking the beginning for some research work on improved performances rather than just applications and performances tuning of machine learning models.

The collected materials from the specified periods of time were, scrutinize, filtered, organized and analyzed after an online search based on the keywords agricultural soil fertility, soil quality, soil health, soil quality assessment, soil fertility analysis, machine learning, deep learning, soil fertility prediction using machine learning techniques, predicting soil fertility using machine learning techniques, agricultural soil data mining, data mining in soil analysis, data science in agricultural soils, agricultural soil analysis in developing countries; from Google Scholar, ResearchGate, Science Direct, IEEE Xplore, Springer Link, Academia.edu, Elsevier, Association for Computing Machinery (ACM), and DBLP databases, as well as from other web-based knowledge sources including machinelarningmastery.com. The materials included conference papers, journal, magazines, newspapers, and encyclopedia articles, books and book chapters, case studies, thesis, patents, reports and documents, manuscripts, dictionary entries, email alerts, instant messages, forum and blog posts, presentations, as well as web pages. Some additional offline materials or hand books of the same were also obtained from libraries and book stores including the Nelson Mandela African Institution of Science and Technology Library; and reviewed. After the three pass method's first phases of quickly scanning and examining the references list to identify relevant research or journal title, the second phase of reading of the potential articles with greater care followed; and the third phase finalized the virtual re-implementation of the articles and aggregation of this survey. Detailed analysis of relevant materials was performed using mixed method of qualitative content in-depth analysis of machine learning applications in agricultural soil nutrients analysis and fertility status prediction and simple quantitative descriptive statistics of the machine learning algorithms and soil parameters use frequency analysis by different researchers, by using Microsoft Excel.

## 4. Machine Learning for Soil Nutrients Analysis and Fertility Status Prediction

The genesis of ML methods in pedology traces back to 1980's, when it was first applied in pedometrics whereby ML data-driven methods could be applied in the modelling and prediction of soil fertility. Presented here is a briefing on a few previous of these works from 2010 to date (2022). These works addressed a range of ML tasks from classifying soil properties into classes of very low, low, moderate, high and very high fertility, to predicting unknown values. Whereas it is best practice to use as many possible algorithms, with all possible available principle parameters in order to perform an exhaustive evaluation so as to attain good analytical results and final model(s). [87] compared the performance of J48, KNN, JRip, NB, SVM, ANN classification algorithms by using PH, EC, N, P, K, OC, S, Fe, Mn,   and Zn input variables of soil dataset to predict soil fertility as 'fertile' or 'not fertile', whereby JRIP scored maximum accuracy of 97%. In another study by [101], data from Vellore soil testing laboratory with soil attributes PH, EC, Fe, Zn, Mn, Cu, OC, P, K, and fertility index (FI) as 'ideal' or 'not ideal' were utilized to perform experiments of training various bagging, boosting, and stacking ensemble classifiers, were they pre-processed the data, extracted relevant features as a means to achieve better performance, and attained an accuracy of 98.15% by boosting the decision tree like C5.0 algorithm. A versatile method for rapid and accurate determination of soil

fertility for sugarcane production was developed in [102], whereby the soil fertility index was established and modelled independently using boosted decision trees with the use of soil attributes PH, OM (OC), Ca,   and Mg, Aluminium used in place of B   due to their study finding high correlation between the two, whereby they achieved AUC scores of 0.76, 0.67 and 0.65 for the respective fertility classes 'highly fertile', 'fertile', or 'least fertile' prediction. In another work, the Random Forest was used to develop a model that was used as part of the work to predict soil's OC, N, P, K, Ca, Mg, Na, Fe, Mn, Cu, Al nutrients fertilities and use the information to understand the edaphic drivers of soil constraints to very extreme high or near zero yields and heterogeneity across Africa, to guide in nutrients-specific interventions, they could find that soil factors could explain 72% of the variations in yields [103]. [104] developed a hybrid classification model by using a Decision Tree Classifier to isolate the soil's PH, EC, OC, N, P, K, S, Zn, Fe, Cu, Mn, and B dependent features and used Naïve Bayes classification on the independent features to predict the fertilities for the primary properties (PH, EC, OC, N) with individual naïve Bayes, and decision tree respective performances of 69.9%, 90.43%, and 99.93% for the DT-NB independent featured hybrid. While they macro P, K, S, Zn, nutrients were respectively predicted at 38%, 88%, 97% accuracies, the micro Fe, Cu, Mn, B nutrients levels were predicted at 42%, 83%, 99.93% accuracies, respectively. [105] examined soil micro and macro nutrients EC, K, pH, Mn, Zn, S, P, B, OC using machine learning to grade soil nutrients, and they applied various classification algorithms and found that random forest had the highest accuracy score as compared to support vector machine and Gaussian naïve Bayes in predicting the soil classes for suitable crop plantation. Likely, [106] used PH, EC, OC, P, K, Fe, Zn, Mn, Cu to implement machine learning models for predicting soil fertility as low, high or medium using Support Vector Machine, nearest neighbor, Naïve Bayes, and Decision Tree that scored 60%. Also, [107] implemented machine learning models for automatically predicting the Indian state of Maharashtra village-wise fertility indices of organic carbon (OC), phosphorus pentoxide (P2O5), iron (Fe), manganese (Mn), and zinc (Zn) by using 76 methods belonging to 20 families including neural networks, deep learning, support vector regression, random forests, partial least squares, bagging and boosting, quantile regression and generalized additive models, among many others. Altogether, as per the Government of India standard fertility levels, the prediction of nutrients fertility indices as low, medium or high achieved the utmost best performance through the ensemble of extremely randomized trees (extraTrees), the results of which corresponded to accuracy (Acc) and Cohen kappa values of (Acc= 86.45% Kappa= 69.60%), (Acc= 79.03% Kappa= 56.19%), (Acc= 79.46% Kappa= 52.51%), (Acc= 86.13% Kappa= 71.08%), (Acc= 97.63% Kappa= 81.03%) for OC, Fe, P2O5, Mn, and Zn, respectively, which is considerably fairly accurate. Other best performing models were those generated through regularized random forest, random forests, and random forest with feature selection, last but not least good performances were obtained from gradient boosting of regression trees (bstTree) and generalized boosting regression (gbm); quantile random forest, M5 rule-based model with corrections based on nearest neighbors (cubist) and support vector regression (svr). In another study, [108] designed an intelligent soil PH, OC, EC, P, K, B nutrient and pH classification using weighted voting ensemble deep learning (ISNpHC-WVE) technique. Such classifications were employed in generating village-wise fertility indices analyses, and they are applied for making fertilizer recommendations using the decision support systems.. In addition, three deep learning (DL) models namely gated recurrent unit (GRU), deep belief network (DBN), and bidirectional long short term memory (BiLSTM) were used for the predictive analysis. Moreover, a weighted voting ensemble model was employed which allows a weight vector on every DL model of the ensemble depending upon the attained accuracy on every class. Furthermore, [109] used different classification algorithms to predict fertility rate based on soil's PH, EC, Fe, Cu, Zn, OC, P, K. Whereby, J48 classifier performed better in predicting fertility index for six (6) classes very low, low, medium, medium high, high, very high with 98.17% accuracy, while naïve bayes and random forest had respective performances of 77.18% ,and 97.92%, their observation generally showed fertility rate for Aurangabad district to be medium. In another study, [110] projected a comparative analysis of Naïve Bayes, JRip and J48 ML algorithms by using soils data with attributes PH, EC, OC, P, K, Fe, Zn, Mn, Cu, it was observed that JRip classification algorithm gave better results compared to the other two algorithms, whereby it

achieved an accuracy of 91.9% and therefore it was recommended to predict six(6) soil classes very high, high, moderately high, moderate, low, and very low. Last but not least, a study by [2] was also useful in providing information on soil features, and algorithms of interest whereby PH, EC, N, OC, P, Ca, Mg, Na, K, Fe, Mn, Cu, and Zn could be observed key features these of which were modelled using naïve Bayes and random forest trees as part of a task to numerically classify a portion of Kilombero Valley soil clusters in Tanzania. Last, but not list, in [111] a novel 2-Stage Hybrid Ensemble Based Heterogeneous Committee Machine for Improving Soil Fertility Status Prediction Performance was developed. Specifically, agricultural soil properties to be attributed as features for soil analysis and fertility prediction by using machine learning algorithms were identified and modelled following a feature selection as OC, pH, EC, TN, P, Ca, K, Mg, Na, S, Mn, Al, Zn, Fe, B. Then machine learning K-Mean clustering algorithm with K-elbow was used to categorize available distinct soil fertility status target classes based crop yields as an index to fertility. Finally heterogeneous hybrid classifiers were evaluated to build a weighted voting ensemble (WVE) with improved prediction performance, by combining the judgments of class probability predictions from the individual hybrid classifiers through optimization in a novel brute based 1EXP(-)Z+ multi precision search spaces for guaranteeing optimality finding. Whereby the K-mean hybrid based WVE combination of GB, RF, SV, KN, DT was the best alternative with accuracy of 98.93% and Cohen Kappa 93.98% on test data, Furthermore, the solution in [111] achieved through ROC analysis AUC score of 0.87, 0.83 and 0.82 for the respective low, medium and high fertility target classes. These results which showed improvement as compared to models in other studies as shown in Table 2 that provides a summary of the reviewed studies related to application of machine learning in soil chemical properties modelling.

**Table 2.** Summary of the State-of-the-art ML-based approaches and Soil chemical properties used in modelling nutrients and fertility status prediction.

| Author | Chemical Properties (features) | Dataset Size | Technique (ML algorithms) | Number of fertilities target classes | Max Accuracy/ROC Performance (%) |
|---|---|---|---|---|---|
| [111] | OC, pH, EC, TN, P, Ca, K, Mg, Na, S, Mn, Al, Zn, Fe, B | 6260 | GB, RF, SVM, KNN, DT | 3 (low, medium, and high) | 98.93%, AUC 0.87, 0.83. 0.82 for respective high, medium, and low fertility classes |
| [101] | PH, EC, Fe, Zn, Mn,Cu, OC, P, K | 1430 | TreeBag and RF ensemble bagging, *boosting C5.0* and boosting Gbm, KNN, CART, SVM, LR via GLM stacking ensemble | 2 (ideal and not ideal) | 98.15 |
| [102] | PH, OC, Ca, Mg, Al | - | Boosted decision trees | 3 (highly fertile, fertile, and least fertile) | 0.76, 0.67, 0.65 for respective high, medium, and low fertility classes |
| [87] | PH, EC, N, P, K, OC, S, Fe, Mn, Zn | 127 | J48, KNN, JRip, NB, SVM, ANN with 10Fold CV and % split | 2 (fertile and not fertile) | 97 |
| [108] | PH, OC,EC, P, K, B | 144 | gated recurrent unit (GRU), deep belief network (DBN), and | low, medium, and high for target classes, pH level | 0.9281% for fertility status, and 0.9497% for PH |

| | | | bidirectional long short term memory (BiLSTM), and WVE | divided into four classes strongly acidic, highly acidic, moderately acidic, and slightly acidic. | |
|---|---|---|---|---|---|
| [105] | EC, K, pH, Mn, Zn, S, P, B, OC | Unspecified | Random Forest Classifier, Support Vector Machine, and Gaussian NB | 3 (Low, medium, and high) | 72.74%, 63.33%, 50.78% |
| [107] | EC, OC, N2 O, P2 O5, Fe, Mn, Zn, and B | 930 | NN, DL, SVR, RF, PLS, bagging and boosting, QR and extraTrees ensemble, Boruta, bstTree and gbm; QRF, cubist and svr. | 3 (Low, medium, and high) per element or compounds | - OC (Acc= 86.45% Kappa= 69.60%) - Fe (Acc= 79.03% Kappa= 56.19%) - P2O5 (Acc= 79.46% Kappa= 52.51%) - Mn (Acc= 86.13% Kappa= 71.08%) - Zn (Acc= 97.63% Kappa= 81.03%) |
| [109] | PH, EC, Fe, Cu, Zn,Mn, OC, P, K | 1639 | J48, Naïve Bayes, Random Forest | 6 (Very low, Low, Medium, Medium high, high, very high) | 98.17, 77.18,97.92 |
| [106] | PH, EC,OC, P, K, Fe,Zn, Mn, Cu | Unidentified | SVM, KNN, Decision Tree, Naïve Bayes | 3 (High, medium , low) | 60% |

## 5. Findings and Discussion

In this Section, findings of this survey are presented and discussed.

*5.1. Identification of Soil Properties for modelling soil fertility status Prediction*

Based on previous research work of machine learning algorithms applications in soil analysis and predictions, a range of algorithms and soil parameters or variables that are used for analysis and prediction could be determined. It could be observed that, with respect to soil parameters the most used includes the chemical properties pH, primary nutrients of nitrogen, phosphorus, and potassium; electrical conductivity, organic carbon, and micro nutrients such as iron, manganese, copper, zinc and boron, soil texture is the mostly used physical properties, while biological properties are often less used, as shown in Figure 4 of the Soil Parameters Use Frequency. And Figure 5 presents the most frequently used ML algorithms in modelling for soil fertility status predictions.
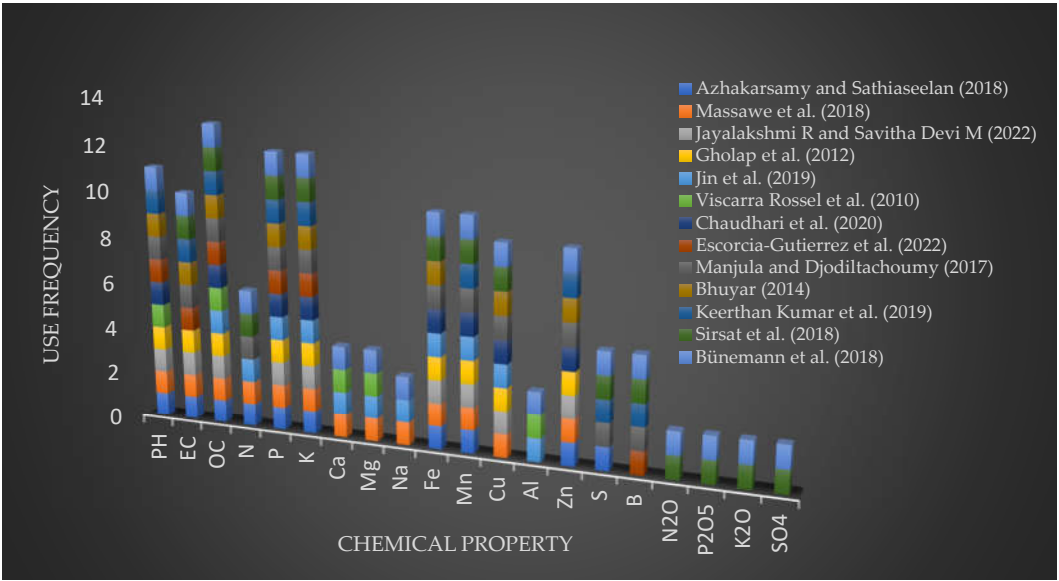
**Figure 4.** Most Frequently used Soil Parameters for modelling for soil fertility status predictions.

*5.2. Identification of Applied Machine Learning Algorithms in Modelling soil fertility status Prediction*

With respect to the machine learning algorithms that were used to model soil nutrients and other chemical properties, as shown in Figure 5, it could be seen that the random forest and Naïve Bayes (NB) were equally predominantly most frequently used. That could be due to the strengths of the RF in merging together a number of individual DTs as described in sub section 1.3.1, and NB probably due to its ability to model even small datasets.
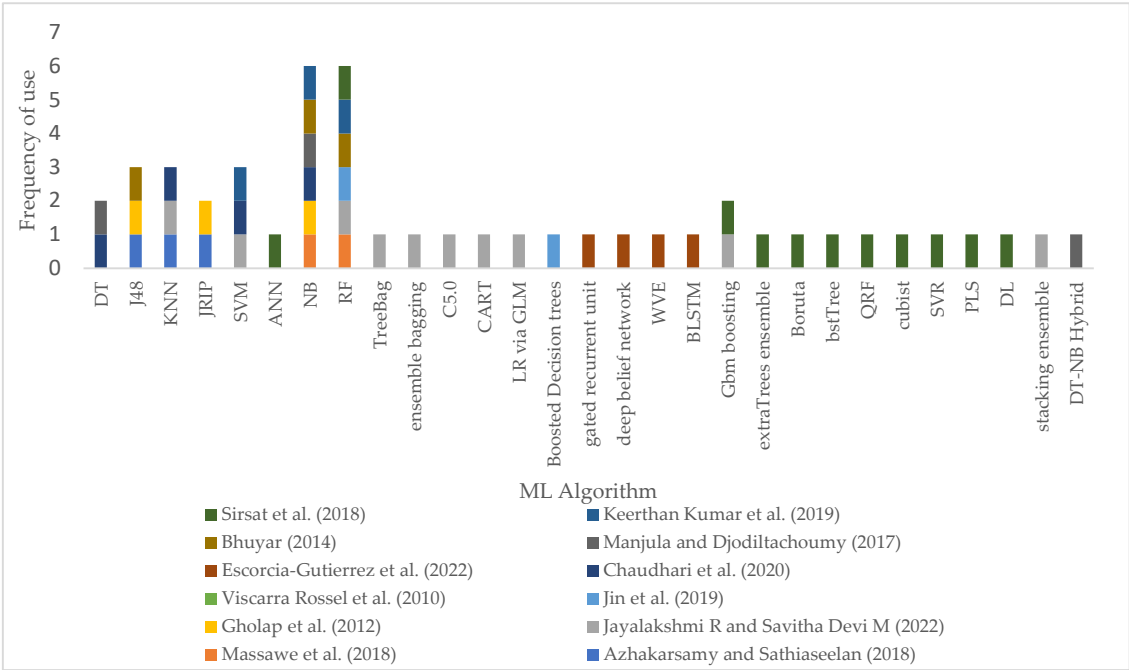


**Figure 5.** Most Frequently used ML Algorithms in modelling for soil fertility status predictions.

SVM, J48 the java version C.5 decision tree, and K-nearest neighbor followed. Next were the decision tree, JRIP, and the gaining popularity gradient boosting algorithms. Unfortunately, the artificial neural network seem to not have been most applied. The main reason could have been the fact that as shown in Table 2 the empirical results, from section 4, most of the dataset used in the

studies seemed not be too large enough to allow for the ANN to converge on smaller datasets which is one of ANN drawbacks highlighted in subsection 1.3.1.

*5.3. Exploited Features, Target Classes, and Reported Model Performances*

5.3.1. Exploited Features and Target Classes

From the empirical literature review, as shown in Table 3, it could be observed that while in classifying soil properties, some studies used only two target classes low and high [87,101]; others used three [102,105,106,108,111], whereas up to six target classes which includes very low, low, medium, medium high, high, very high [109]. This of which could have positively or negatively impact site specific fine-tuned or optimal fertility restorations or preservation dosages. Whereby the actual fact of that impact could be determined only if field plantation experimentation would be conducted for validating the predictions provided thereof. Such experiments which are so far not disclosed by the involved researchers. Whereby three fertility class targets which were derived by using the K-means clustering algorithms with Knee detection method to determine the optimal number of clusters by [111] in another field maize plantation experimentation observe the impact of classification models predictions delivered and used as a guide to recommend for soil treatment using three number of optimal clusters as a representation of soil fertility class targets.

**Table 3.** Exploited features with varied target classes.

| Author | Number of Features | Number of Target Classes |
|---|---|---|
| [111] | 15 | 3 |
| [101] | 9 | 2 |
| [102] | 5 | 3 |
| [87] | 10 | 2 |
| [108] | 6 | 3 |
| [105] | 9 | 3 |
| [109] | 9 | 6 |
| [106] | 9 | 3 |

5.3.2. Reported Model Performances

As shown in Table 4, the models AUC-ROC performances in continuously providing for positive predictions of either of the soil fertility class targets using three fertility status levels low, medium and high seemed to be affected and vary with the number of features used, with the corresponding Plot of AUC-ROC Performances based on number of features portrayed in Figure 6 (a). This could deduced from observing studies by [111] and [102], which used similar number of class targets but reported different performance. Although, the context of the data being from two far different geographical location (South America in 102, and Africa in [111]) with different climatic and weather conditions might have accounted for such model performance variation, other than the number of features being different. This survey however, may overrule that assertion due the fact that the researcher or modeler in [111] use of the weighted voting ensemble (WVE) of heterogeneous classifiers technique for performance improvement could have been the reason behind such an improved performance with AUC-ROC of 0.82, 0.83, and 0.87 as compared to [102], which was observed to only use homogeneous boosted decision trees to achieve AUC-ROC of 0.65, 0.67, and 0.76 for respective fertility class targets low, medium, and high.

Whereas, the predictive performance was seen to be affected by not only the number of features used to model the fertility prediction classifiers, but also as shown in Figure 6 (a) and (b) by: The predictive accuracies were seen to be highly impacted by the modelling technique, whereby ensemble techniques showed better performance over the individual algorithm based models. And also, the ensemble diversity which was observantly significant in influencing the model performance, whereby more heterogeneous model members based ensemble in [111] that consisted of the GB, RF, SV, KNN, and DT WVE combination as shown in Table 2 in Section 4, slightly presented better

performances than the homogenous counterparts in [101] created by boosting individual C5.0 models.

**Table 4.** AUC-ROC Performances based on number of features.

| Author | Number of Features | ROC-Low | ROC-Medium | ROC-High |
|--------|-------------------|---------|------------|----------|
| [111] | 15 | 0.82 | 0.83 | 0.87 |
| [102] | 5 | 0.65 | 0.67 | 0.76 |



(**a**)                                              (**b**)
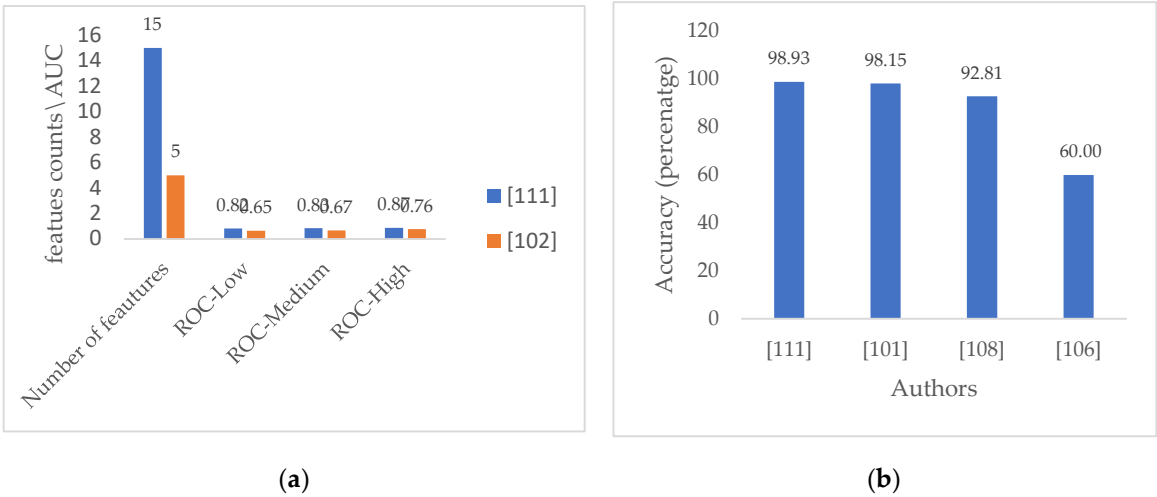
**Figure 6.** (a) Plot of AUC-ROC Performances based on number of features; (b) Variations in reported soil fertility prediction model performances.

## 6. Conclusions and Recommendations

In this paper a survey of machine learning modelling for soil nutrients and other chemical properties analysis and soil fertility status prediction was presented. Empirical results showed that while the chemical properties potential Hydrogen (pH) is frequently used, the organic carbon, phosphorus, potassium, followed by iron, manganese, copper and zinc are the most frequently used soil nutrients for soil fertility prediction. With respect to machine learning modelling algorithms used to model these features, the Random Forest and Naïve Bayes were predominant as they showed high frequency in use, Followed by Support Vector Machine, K-Nearest Neighbors and J48, while Decision Trees, JRIP, and the Gradient Boosting which is now gaining popularity due to its proven abilities to fit data and model high performance models, were next in frequency of use. The survey could draw from various studies the use of a range of soil fertility target classes starting from two, three, up to six classes used targets for fertility status, with a study by [111] heuristically suggesting for three target classes as optimal for representation of fertility status. Finally, a varying range of model performances were observed from previous studies, with the highest reported accuracy being 98.93%, followed by 98.15%, and the model with the least performance had achieved only 60% prediction performance. While, the use of an ensemble of heterogeneous classifiers models for improving model predictive performance became slightly superior over homogenous ensemble,  it maintains the theoretical fact that ensemble with diverse members is favorable in the significant improvement of model performance. This fact which is cemented by the empirical findings ascertained in Figure 6 (b) of the reported soil fertility prediction model performances, whereby, it could be seen that although the homogenous ensemble in study in [101] attained a very good predictive accuracy of 98.15%, but was observably to be slightly outperformed by the heterogeneous ensemble in [111] that scored 98.93%.

While, the findings in this survey shows good promises to provide farmers and other relevant agricultural stakeholders with models that provide optimal analytical information and

recommendations for use in decision making through fine-tuned site specific treatments and management. As recommendations for further research and developments related to machine learning modelling for agricultural soil properties analysis and fertility status Predictions and use of machine learning in general: as there lacks studies focusing on implementation of ensemble methods for predicting soil fertility status by using soil nutrients and chemical properties data. Giving more considerations in their implementation could beneficial in delivering models with high performance as it could be seen from the results of the few reported studies. Also, inclusion of climatic and weather conditions data along with soil nutrients and chemical properties soil fertility could be considered in order to provide for a more feature exhaustive model which may addresses more characteristics that are necessary modelling a more comprehensive soil fertility ecosystem. Furthermore, researchers or machine learning engineers of the classical machine learning dependent on a prior exhaustive evaluation as with candidate algorithm model optimization through parameters tuning may also highly consider on improving predictive performances through ensembles such as WVE schemes of diverse models, which these in turn are subjected to optimization, in order to effectively improve model prediction performance, such as these for soil nutrients analysis and fertility status so as to manage agricultural soil fertility more precisely. Last but not least, machine learning related research and developments such applications in the more digitally artificial intelligent smart agricultural soil fertility under lacking systems regions may help in bridging the agricultural productivity gap and raise the agricultural efforts to corresponding gross domestic products ratios. In turn, this may speed up the making of a global food productivity and supply system become a reality, through smart fertility management systems which are one of the key fundamental factors necessary to improve crop growth and productivity. This which may lead to sustainability in in the insurance of food security for the rapid world's population growth based on the observed Food and Agriculture Organization (FAO) population growth estimates that may come into sight. And most critically if carefully implemented in the most food under lacking regions it may deal with the current food shortages in those regions.

## References

1.  Manjula, E.; Djodiltachoumy, S. Data Mining Technique to Analyze Soil Nutrients Based on Hybrid Classification. *IJARCS* **2017**. https://doi.org/DOI: http://dx.doi.org/10.26483/ijarcs.v8i8.4794.
2.  Massawe, B. H.; Subburayalu, S. K.; Kaaya, A. K.; Winowiecki, L.; Slater, B. K. Mapping Numerically Classified Soil Taxa in Kilombero Valley, Tanzania Using Machine Learning. *Geoderma* **2018**, *311*, 143–148.
3.  Ferrández-Pastor, F. J.; García-Chamizo, J. M.; Nieto-Hidalgo, M.; Mora-Martínez, J. Precision Agriculture Design Method Using a Distributed Computing Architecture on Internet of Things Context. *Sensors* **2018**, *18* (6), 1731.
4.  Bünemann, E. K.; Bongiorno, G.; Bai, Z.; Creamer, R. E.; De Deyn, G.; de Goede, R.; Fleskens, L.; Geissen, V.; Kuyper, T. W.; Mäder, P. Soil Quality–A Critical Review. *Soil Biology and Biochemistry* **2018**, *120*, 105–125.
5.  Cherlinka, V. *Soil Fertility: Influencing Factors And Improvement Strategies*. https://eos.com/blog/soil-fertility/ (accessed 2023-08-09).

6.    FAO. *The Future of Food and Agriculture – Trends and Challenges*; Food and Agriculture Organization of the United Nations: Rome, 2017.

7.    FAO. Agricultural Outlook 2016-2025., 2016.

8.    Ishengoma, F.; Athuman, M. Internet of Things to Improve Agriculture in Sub Sahara Africa-a Case Study. **2018**.

9.    Jayaraman, P.; Yavari, A.; Georgakopoulos, D.; Morshed, A.; Arkayd, Z. Internet of Things Platform for Smart Farming: Experiences and Lessons Learnt. **2016**.

10.   Sirsat, M. S.; Cernadas, E. G.; Delgado, M. F.; Khan, R. Classification of Agricultural Soil Parameters in India. *Computers and electronics in agriculture* **2017**, *135*, 269–279.

11.   Sirsat, M. S.; Cernadas, E. G.; Delgado, M. F. Application of Machine Learning to Agricultural Soil Data. PhD Thesis, Universidade de Santiago de Compostela, 2017.

12.   Vijayabaskar, P. S.; Sreemathi, R.; Keertanaa, E. Crop Prediction Using Predictive Analytics. In *2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*; IEEE, 2017; pp 370–373.

13.   Kavvadias, V.; Papadopoulou, M.; Vavoulidou, E.; Theocharopoulos, S.; Malliaraki, S.; Agelaki, K.; Koubouris, G.; Psarras, G. Effects of Carbon Inputs on Chemical and Microbial Properties of Soil in Irrigated and Rainfed Olive Groves. In *Soil Management and Climate Change*; Elsevier, 2018; pp 137–150.

14.   Gholap, J.; Ingole, A.; Gohil, J.; Gargade, S.; Attar, V. Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction. *arXiv preprint arXiv:1206.1557* **2012**.

15.   Liu, Y.; Wang, H.; Zhang, H.; Liber, K. A Comprehensive Support Vector Machine-Based Classification Model for Soil Quality Assessment. *Soil and Tillage Research* **2016**, *155*, 19–26.

16.   Emerson, S.; Cranston, R. E.; Liss, P. S. Redox Species in a Reducing Fjord: Equilibrium and Kinetic Considerations. *Deep Sea Research Part A. Oceanographic Research Papers* **1979**, *26* (8), 859–878.

17.   Kommineni, M.; Perla, S.; Yedla, D. B. A Survey of Using Data Mining Techniques for Soil Fertility. **2018**.

18.   Rajeswari, V.; Arunesh, K. Analysing Soil Data Using Data Mining Classification Techniques. *Indian journal of science and Technology* **2016**, *9* (19), 1–4.

19.   Yusof, K. M.; Isaak, S.; Rashid, N. C. A.; Ngajikin, N. H.; Bahru, U. J. NPK Detection Spectroscopy on Non-Agriculture Soil. *Jurnal Teknologi* **2016**, *78* (11), 227–231.

20.   Kamilaris, A.; Kartakoullis, A.; Prenafeta-Boldú, F. X. A Review on the Practice of Big Data Analysis in Agriculture. *Computers and Electronics in Agriculture* **2017**, *143*, 23–37.

21.   Sharma, A.; Weindorf, D. C.; Wang, D.; Chakraborty, S. Characterizing Soils via Portable X-Ray Fluorescence Spectrometer: 4. Cation Exchange Capacity (CEC). *Geoderma* **2015**, *239*, 130–134.

22.   Walsh, M.; Meliyo, J.; Wu, W.; Chen, J.; Shepherd, K.; Ekise, C.; Simbila, W.; Sila, A. M.; Zhan, Y.; Mulvey, J. Tanzania Soil Information Service (TanSIS). **2018**. https://doi.org/10.17605/OSF.IO/4NGAU.

23.   Alpaydin, E. *Introduction to Machine Learning*; MIT press, 2020.

24.   Baştanlar, Y.; Özuysal, M. Introduction to Machine Learning. *miRNomics: MicroRNA biology and computational analysis* **2014**, 105–128.

25.   Bonaccorso, G. *Machine Learning Algorithms*; Packt Publishing Ltd, 2017.

26.   Hurwitz, J.; Kirsch, D. *Machine Learning for Dummies*, IBM limited edition.; John Wiley and Sons, Inc.: 111 River St. Hoboken, NJ 07030-5774, 2018.

27.   McQueen, R. J.; Garner, S. R.; Nevill-Manning, C. G.; Witten, I. H. Applying Machine Learning to Agricultural Data. *Computers and electronics in agriculture* **1995**, *12* (4), 275–293.

28.   Mishra, S.; Mishra, D.; Santra, G. H. Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper. *Indian Journal of Science and Technology* **2016**, *9* (38).

29.   Mitchell, T. *Machine Learning*; McGraw Hill: Ithaca, NY, 1997.

30.   Ayodele, T. O. Types of Machine Learning Algorithms. In *New advances in machine learning*; IntechOpen, 2010.

31.   Anifowose, F. *Hybrid Machine Learning Explained in Nontechnical Terms*. JPT. https://jpt.spe.org/hybrid-machine-learning-explained-nontechnical-terms (accessed 2022-10-09).

32.   Nilsson, N. J. Introduction to Machine Learning (Draft Version), Department of Computer Science, 1999. https://ai.stanford.edu/~nilsson/MLbook.pdf.

33.   Golge, E. *Brief History of Machine Learning*. A Blog From a Human-engineer-being. http://www.erogol.com/brief-history-machine-learning/.

34.  Chaurasia, V.; Pal, S. Performance Analysis of Data Mining Algorithms for Diagnosis and Prediction of Heart and Breast Cancer Disease. **2017**.

35.  Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Drug Design by Machine Learning: Support Vector Machines for Pharmaceutical Data Analysis. *Computers & chemistry* **2001**, *26* (1), 5–14.

36.  Zhang, L.; Tan, J.; Han, D.; Zhu, H. From Machine Learning to Deep Learning: Progress in Machine Intelligence for Rational Drug Discovery. *Drug discovery today* **2017**, *22* (11), 1680–1685.

37.  Adhatrao, K.; Gaykar, A.; Dhawan, A.; Jha, R.; Honrao, V. Predicting Students' Performance Using ID3 and C4. 5 Classification Algorithms. *arXiv preprint arXiv:1310.2071* **2013**.

38.  Durairaj, M.; Vijitha, C. Educational Data Mining for Prediction of Student Performance Using Clustering Algorithms. *International Journal of Computer Science and Information Technologies* **2014**, *5* (4), 5987–5991.

39.  Ameri, S.; Fard, M. J.; Chinnam, R. B.; Reddy, C. K. Survival Analysis Based Framework for Early Prediction of Student Dropouts. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*; ACM, 2016; pp 903–912.

40.  Mduma, N.; Kalegele, K.; Machuve, D. A Survey of Machine Learning Approaches and Techniques for Student Dropout Prediction. *Data Science Journal* **2019**, *18* (1).

41.  Brownlee, J. *Data Mining with Weka*; 2016.

42.  Castle, N. *An Introduction to Machine Learning Algorithms*. Oracle + Datascience.

43.  Kamilaris, A.; Prenafeta-Boldú, F. X. Deep Learning in Agriculture: A Survey. *Computers and Electronics in Agriculture* **2018**, *147*, 70–90.

44.  LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *nature* **2015**, *521* (7553), 436.

45.  Masri, D.; Woon, W. L.; Aung, Z. Soil Property Prediction: An Extreme Learning Machine Approach. In *International Conference on Neural Information Processing*; Springer, 2015; pp 18–27.

46.  Sathya, R.; Abraham, A. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence* **2013**, *2* (2), 34–38.

47.  Bhattacharya, B.; Solomatine, D. P. Machine Learning in Soil Classification. *Neural networks* **2006**, *19* (2), 186–195.

48.  Domingos, P. A Few Useful Things to Know about Machine Learning. *Communications of the ACM* **2012**, *55* (10), 78–87.

49.  Jadhav, S. D.; Channe, H. P. Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research* **2016**, *5* (1).

50.  Quinlan, J. R. Induction of Decision Trees. *Machine learning* **1986**, *1* (1), 81–106.

51.  Gholap, J. Performance Tuning of J48 Algorithm for Prediction of Soil Fertility. *arXiv preprint arXiv:1208.3943* **2012**.

52.  Melville, P.; Mooney, R. J. Constructing Diverse Classifier Ensembles Using Artificial Training Examples. In *IJCAI*; 2003; Vol. 3, pp 505–510.

53.  Melville, P.; Shah, N.; Mihalkova, L.; Mooney, R. J. Experiments on Ensembles with Missing and Noisy Data. In *International Workshop on Multiple Classifier Systems*; Springer, 2004; pp 293–302.

54.  Pastur-Romay, L. A.; Cedrón, F.; Pazos, A.; Porto-Pazos, A. B. Deep Artificial Neural Networks and Neuromorphic Chips for Big Data Analysis: Pharmaceutical and Bioinformatics Applications. *International journal of molecular sciences* **2016**, *17* (8), 1313.

55.  Siegelmann, H.; Sontag, E. Computational Power of Neural Networks. *Journal of Computer and System Sciences* **1995**, *50* (1), 132–150.

56.  Yedjour, D.; Benyettou, A. Symbolic Interpretation of Artificial Neural Networks Based on Multiobjective Genetic Algorithms and Association Rules Mining. *Applied Soft Computing* **2018**, *72*, 177–188.

57.  Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning Representations by Back-Propagating Errors. *Cognitive modeling* **1988**, *5* (3), 1.

58.  Hochreiter, S. Investigations on Dynamic Neural Networks. *Diploma, Technische Universität München* **1991**, *91* (1).

59.  Hochreiter, S. Studies on Dynamic Neural Networks [in German] Diploma Thesis. *TU Münich* **1991**.

60.  Cortes, C.; Vapnik, V. Support-Vector Networks. *Machine Learning* **1995**, *20* (3), 273–297. https://doi.org/10.1007/BF00994018.

61.  Vapnik, V.; S, G.; A, S. *Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing*, M. Mozer, M. Jordan, and T. Petsche.; Advances in Neural Information Processing Systems; MIT Press: Cambridge, MA, 1997.

62. Poorinmohammad, N.; Mohabatkar, H.; Behbahani, M.; Biria, D. Computational Prediction of Anti HIV-1 Peptides and in Vitro Evaluation of Anti HIV-1 Activity of HIV-1 P24-Derived Peptides. *Journal of Peptide Science* **2015**, *21* (1), 10–16.

63. Breiman, L. Random Forests. *Machine learning* **2001**, *45* (1), 5–32.

64. Ho, T. K. Random Decision Forests. *Proceedings of the Third International Conference on Document Analysis and Recognition* **1995**, *1*, 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

65. Marr, B. A Short History of Machine Learning – Every Manager Should Read. *Forbes* **2016**.

66. Badillo, S.; Banfai, B.; Birzele, F.; Davydov, I. I.; Hutchinson, L.; Kam-Thong, T.; Siebourg-Polster, J.; Steiert, B.; Zhang, J. D. An Introduction to Machine Learning. *Clinical pharmacology & therapeutics* **2020**, *107* (4), 871–885.

67. Bengio, Y. Learning Deep Architectures for AI. *Foundations and trends® in Machine Learning* **2009**, *2* (1), 1–127.

68. LeCun, Y.; Bengio, Y. Convolutional Networks for Images, Speech, and Time Series. *The handbook of brain theory and neural networks* **1995**, *3361* (10), 1995.

69. Pearce, J.; Ferrier, S. Evaluating the Predictive Performance of Habitat Models Developed Using Logistic Regression. *Ecological modelling* **2000**, *133* (3), 225–245.

70. Hanley, J. A. Receiver Operating Characteristic (ROC) Methodology: The State of the Art. *Crit Rev Diagn Imaging* **1989**, *29* (3), 307–335.

71. Obuchowski, N. A.; Bullen, J. A. Receiver Operating Characteristic (ROC) Curves: Review of Methods with Applications in Diagnostic Medicine. *Physics in Medicine & Biology* **2018**, *63* (7), 07TR01.

72. Brownlee, J. *How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification*. Machine Learning Mastery. https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/ (accessed 2022-09-24).

73. Soleymani, R.; Granger, E.; Fumera, G. F-Measure Curves: A Tool to Visualize Classifier Performance under Imbalance. *Pattern Recognition* **2020**, *100*, 107146.

74. Hall, D.; McCool, C.; Dayoub, F.; Sunderhauf, N.; Upcroft, B. Evaluation of Features for Leaf Classification in Challenging Conditions. In *2015 IEEE Winter Conference on Applications of Computer Vision*; IEEE, 2015; pp 797–804.

75. Osisanwo, F. Y.; Akinsola, J. E. T.; Awodele, O.; Hinmikaiye, J. O.; Olakanmi, O.; Akinjobi, J. Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology (IJCTT)* **2017**, *48* (3), 128–138.

76. Witten, I. H.; Frank, E.; Hall, M. A.; Pal, C. J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann, 2016.

77. Solutions, E. *Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures*. Exsilio Blog. https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/ (accessed 2022-09-24).

78. Minh, D. H. T.; Ienco, D.; Gaetano, R.; Lalande, N.; Ndikumana, E.; Osman, F.; Maurel, P. Deep Recurrent Neural Networks for Mapping Winter Vegetation Quality Coverage via Multi-Temporal SAR Sentinel-1. *arXiv preprint arXiv:1708.03694* **2017**.

79. Kumar, A. *Cohen Kappa Score Python Example: Machine Learning*. Data Analytics. https://vitalflux.com/cohen-kappa-score-python-example-machine-learning/ (accessed 2022-09-02).

80. *Cohen's Kappa: Learn It, Use It, Judge It*. KNIME. https://www.knime.com/blog/cohens-kappa-an-overview (accessed 2022-09-24).

81. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *Journal of machine learning research* **2011**, *12* (Oct), 2825–2830.

82. Michalski, R. S. Learning by Being Told and Learning from Examples: An Experimental Comparison of the Two Methods of Knowledge Acquisition in the Context of Development an Expert System for Soybean Disease Diagnosis. *International Journal of Policy Analysis and Information Systems* **1980**, *4* (2), 125–161.

83. Bagheri Bodaghabadi, M.; Moghimi, A.; Navidi, M. N.; Ebrahimi Meymand, F. Modeling of Yield and Rating of Land Characteristics for Corn Based on Artificial Neural Network and Regression Models in Southern Iran. *Desert* **2018**, *23* (1), 85–95.

84. Devi, M. P. K.; Anthiyur, U.; Shenbagavadivu, M. S. Enhanced Crop Yield Prediction and Soil Data Analysis Using Data Mining. *International Journal of Modern Computer Science* **2016**, *4* (6).

85. Negied, N. K. Expert System for Wheat Yields Protection in Egypt (ESWYP). *International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN* **2014**, 2278–3075.

86. Saini, H. S.; Kamal, R.; Sharma, A. N. Web Based Fuzzy Expert System for Integrated Pest Management in Soybean. *International Journal of Information Technology* **2002**, *8* (1), 55–74.

87. Azhakarsamy, R.; Sathiaseelan, J. G. R. *Comparison of Classifiers to Predict Classification Accuracy for Soil Fertility*; SSRN Scholarly Paper ID 3315395; Social Science Research Network: Rochester, NY, 2018. https://papers.ssrn.com/abstract=3315395 (accessed 2019-04-12).

88. Ramane, D. V.; Patil, S. S.; Shaligram, A. D. Detection of NPK Nutrients of Soil Using Fiber Optic Sensor. In *International Journal of Research in Advent Technology Special Issue National Conference ACGT 2015*; 2015; pp 13–14.

89. Adamchuk, V. I.; Hummel, J. W.; Morgan, M. T.; Upadhyaya, S. K. On-the-Go Soil Sensors for Precision Agriculture. *Computers and Electronics in Agriculture* **2004**, *44* (1), 71–91. https://doi.org/10.1016/j.compag.2004.03.002.

90. Krupicka, J.; Sarec, P.; Novak, P. Measurement of Electrical Conductivity of Fertilizer NPK 20-8-8. In *proceedings of the international scientific conference*; Latvia University of Agriculture, 2016.

91. Kumar, A.; Kannathasan, N. A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining. *IJCSI International Journal of Computer Science Issues* **2011**, *8* (3).

92. CIA Factbook. *Labor force - by occupation*. https://www.indexmundi.com/factbook/fields/labor-force-by-occupation (accessed 2019-04-29).

93. CIA Factbook. *GDP - composition by sector*. https://www.indexmundi.com/factbook/fields/gdp-composition-by-sector (accessed 2019-04-29).

94. Haq, K.; Officer, G. E. Managing and Reviewing the Literature. *Strategies* **2014**, *9*, 10–15.

95. Keshav, A.; Cheriton, D. R. How to Read a Paper. **2013**, 2.

96. Sultan, M.; Miranskyy, A. Ordering Stakeholder Viewpoint Concerns for Holistic and Incremental Enterprise Architecture: The W6H Framework. *arXiv preprint arXiv:1509.07360* **2015**.

97. Sultan, M.; Miranskyy, A. Ordering Interrogative Questions for Effective Requirements Engineering: The W6H Pattern. In *2015 IEEE Fifth International Workshop on Requirements Patterns (RePa)*; IEEE, 2015; pp 1–8.

98. Sultan, M.; Miranskyy, A. Ordering Stakeholder Viewpoint Concerns for Holistic Enterprise Architecture: The W6H Framework. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*; ACM, 2018; pp 78–85.

99. Hebb, D. O. *The Organization of Behaviour New York*; Wiley & Sons, 1949.

100. Markoff, J. BUSINESS TECHNOLOGY; What's the Best Answer? It's Survival of the Fittest. *New York Times* **1990**.

101. Jayalakshmi R; Savitha Devi M. Mining Agricultural Data to Predict Soil Fertility Using Ensemble Boosting Algorithm. *International Journal of Information Communication Technologies and Human Development (IJICTHD)* **2022**, *14* (1), 1–10. https://doi.org/10.4018/IJICTHD.299414.

102. Viscarra Rossel, R. A.; Rizzo, R.; Demattê, J. A. M.; Behrens, T. Spatial Modeling of a Soil Fertility Index Using Visible–Near-Infrared Spectra and Terrain Attributes. *Soil science society of America journal* **2010**, *74* (4), 1293–1300.

103. Jin, Z.; Azzari, G.; You, C.; Di Tommaso, S.; Aston, S.; Burke, M.; Lobell, D. B. Smallholder Maize Area and Yield Mapping at National Scales with Google Earth Engine. *Remote sensing of environment* **2019**, *228*, 115–128.

104. Manjula, E.; Djodiltachoumy, S. Data Mining Technique to Analyze Soil Nutrients Based on Hybrid Classification. *International Journal of Advanced Research in Computer Science* **2017**, *8* (8).

105. Keerthan Kumar, T. G.; Shubha, C.; Sushma, S. A. Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine Learning. *Int J Innov Technol Explor Eng* **2019**, *9* (1), 1301–1304.

106. Chaudhari, R.; Chaudhari, S.; Shaikh, A.; Chiloba, R.; Khadtare, T. D. Soil Fertility Prediction Using Data Mining Techniques. *Bulletin Monumental* **2020**, *21* (01), 8.

107. Sirsat, M. S.; Cernadas, E.; Fernández-Delgado, M.; Barro, S. Automatic Prediction of Village-Wise Soil Fertility for Several Nutrients in India Using a Wide Range of Regression Methods. *Computers and electronics in agriculture* **2018**, *154*, 120–133.

108. Escorcia-Gutierrez, J.; Gamarra, M.; Soto-Diaz, R.; Pérez, M.; Madera, N.; F. Mansour, R. Intelligent Agricultural Modelling of Soil Nutrients and PH Classification Using Ensemble Deep Learning Techniques. **2022**.

109. Bhuyar, V. Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District. *Int. J. Emerg. Trends Technol. Comput. Sci* **2014**, *3* (2), 200–203.

110. Gholap, J.; Ingole, A.; Gohil, J.; Gargade, S.; Attar, V. Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction. *arXiv preprint arXiv:1206.1557* **2012**.

111. Malamsha, A. J.; Dida, M. A.; Moebs, S. 2-Stage Hybrid Based Heterogeneous Ensemble Committee Machine for Improving Soil Fertility Status Prediction Performance. In *International Conference for Technological Advancement in Embedded and Mobile Systems*; ICTA-EMOS, 2023.