Agricultural Water Management 294 (2024) 108710

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat

Application of machine learning approaches in supporting irrigation decision making: A review

Lisa Umutoni, Vidya Samadi *

Department of Agricultural Sciences, Clemson University, Clemson, SC, USA

ARTICLE INFO

Handling editor: Dr Z Xiying

Keywords:
Machine Learning
Irrigation Water Use Modeling
Water Management
FAIR data

ABSTRACT

Irrigation decision-making has evolved from solely depending on farmers' decisions taken based on the visual analysis of field conditions to making decisions based on crop water need predictions generated using machine learning (ML) techniques. This paper reviews ML related articles to discuss how ML has been used to enhance irrigation decision making. We reviewed 16 studies that used ML approaches for irrigation scheduling prediction and decision-making focusing on the input features, algorithms used and their applicability in real world conditions. ML performances in terms of accuracy, water conservation compared to fixed or threshold-based methods are discussed along with modeling performances. Informed by the 16 research studies, we assessed constraints to the adoption of ML in irrigation decision making at field scale, which include limited data availability coupled with data sharing constraints, and a lack of uncertainty quantification as well as the need for physics informed ML based irrigation scheduling models. To address these limitations, we discussed approaches in future research such as integrating process-based models with ML, incorporating expert knowledge into the modeling procedure, and making data and tools Findable, Accessible, Interoperable, and Reusable (FAIR). These approaches will improve ML modeling outcomes and boost the availability of farm-related data and tools for FAIRer data-driven applications of irrigation modeling.

1. Introduction

Irrigation scheduling is becoming an increasingly crucial decision-making task whose goal is to achieve effective and efficient use of water (Saggi and Jain, 2022). The crop quality and yield are significantly dependent on the amount of water and timing. The objective of irrigation scheduling is to apply an adequate amount of water at the right time to a specific crop. However, according to a survey by the US Department of Agriculture, more than 75% of irrigation scheduling methods applied by farmers in the US are based on the Checkbook method (Vellidis et al., 2016), and the condition of crops such as visual observations, crop calendars and observing what the neighbors are doing (USDA, 2017). Inefficient irrigation scheduling methods may result in over-irrigation or under-irrigation. Consequently, water scarcity, nutrient leaching, and increased soil salinity if groundwater is used for irrigation can be observed in over-irrigated regions or yield an economic loss in under-irrigated farms. Science- and technology-based irrigation scheduling approaches, as opposed to the Checkbook method or the condition of crop method, may boost crop profit while reducing environmental consequences by limiting crop water stress (Zhang et al., 2021).

Model-based crop water demand assessment falls under three categories, deterministic methods, process-based models, and machine learning (ML) methods. Deterministic methods such as FAO-56 calculate irrigation water demand based on crop evapotranspiration approaches (Allen et al., 2005). Process-based models such as Soil and Water Assessment Tool (SWAT), Aquacrop, Decision Support System for Agrotechnology Transfer (DSSAT), and the Agricultural Production Systems sIMulator (APSIM) usually determine crop water requirement by simulating biophysical processes leading to plants' growth in the soil-plant-atmosphere system using mathematical formulations of those processes. ML models, on the other hand, are data-driven, they estimate irrigation needs by learning the hidden function that relates input weather and soil data to crop water demands. All techniques have been applied in irrigation scheduling for different crops in various regions across the globe. However, process-based models are highly sensitive to uncertainties associated with physical processes such as initial and boundary conditions and spatiotemporal variability of soil moisture

* Corresponding author.
E-mail address: samadi@clemson.edu (V. Samadi).

https://doi.org/10.1016/j.agwat.2024.108710
Received 29 June 2023; Received in revised form 22 January 2024; Accepted 29 January 2024
Available online 9 February 2024

conditions than are deterministic and ML models (Karandish and Šimůnek, 2016; Gumiere et al., 2020; Li et al., 2020). Consequently, process-based models need extensive and constant calibration as new data are generated to accurately compute irrigation needs.

Furthermore, process-based models are more suitable for irrigation planning than for real-time irrigation decision-making (Gu et al., 2020). Process-based models lack a way to explicitly enhance real-time soil moisture prediction by incorporating weather forecasts or feedback from soil moisture sensors. For instance, SWAT has shown consistent overestimation of the irrigation amount when used for irrigation scheduling which compromises the model's accuracy in automated irrigation control (Sun and Ren, 2014; Maier and Dietrich, 2016). Fortunately, real-time irrigation prediction and scheduling can be achieved thanks to technological developments such as Wireless Sensor Networks (WSN) that enhance data availability and the increased computational power enabling the operation of ML models in real-time. Examples of such applications include an irrigation system developed by Glória et al. (2021) to attain the best water-saving schedule using real-time data, and smart irrigation strategies that predict irrigation water use using several ML algorithms (see Tace et al., 2022).

The application of ML in irrigation scheduling promoted agricultural productivity and water use efficiency (Hunsaker et al., 2005; Karasekreter et al., 2013; Nachankar et al., 2018; Nawandar and Satpute, 2019; Jamroen et al., 2020; Jimenez et al., 2020). With novel data collection techniques, the quantity of available data at a finer spatial-temporal scale, also known as big data, has increased in an unprecedented way (Donratanapat et al., 2020; Huang et al., 2019). Big data can be leveraged to find appropriate solutions for the problems faced by farmers (García et al., 2020), such as water shortage, and crop and cost management (Elijah et al., 2018). ML technologies can be leveraged to use big data for irrigation water use simulation. Indeed, the integration of ML technologies with big data enables irrigation estimation to be undertaken in more intelligent and efficient ways hence ensuring water sustainability for the growing world's population (Sharma et al., 2021).

Recent studies have investigated how ML can benefit the agricultural sectors by improving irrigation systems (Nemali and van Iersel, 2006; Liu et al., 2021). ML has been used to improve traditional ways of carrying out agricultural activities such as irrigation control and management (Tseng et al., 2018; Dahane et al., 2020; Torres-Sanchez et al., 2020). Indeed, ML has been employed to build intelligent data-driven models and provide the best solution for irrigation scheduling by learning the hidden relationship in datasets. The outcomes of the ML irrigation model can benefit farmers in numerous ways. ML algorithms can predict irrigation water requirements based on the study of evaporation processes through collected data and determine soil moisture variations over time. The soil moisture content can then be used as an indicator of when and how much to irrigate. This information can provide early intelligence to farmers to enhance their irrigation decisions.

Currently, there are several ML methods available to predict irrigation water demand that have been applied in different farm settings. Liou et al. (2001) were among the first scholars who applied ML models, i.e., an error-propagation learning back-propagation neural network, to estimate soil moisture from simulated brightness temperature. In a sequence, Dibike et al. (2001) demonstrated the potential of support vector machine (SVM) to outperform artificial neural network (ANN) in remote sensing image classification and regression problems. Gill et al. (2006) used SVM to predict soil moisture in four to seven days ahead of time using historical data with forecasts that were quite comparable to the actual soil moisture measurements. In another study, Shrestha and Shukla (2015) utilized SVM to estimate crop evapotranspiration using lysimeter-based measurements for vine and erect crops to train the model. They found that the SVM model surpassed ANN and boosted the accuracy of FAO-56 crop evapotranspiration estimates by 4%. Soil moisture measurements became possible with the emergence of the Internet of Things (IoT) and the availability of enhanced and more accessible soil sensing and field communication technology (Kamienski et al., 2018; García et al., 2020).

Major drawbacks of ML algorithms, however, are massive data requirement, black-boxiness, uncertainty associated with simulation (Tabas and Samadi, 2022), automatically deducing the features, and optimally tuning them for the desired outcomes (e.g., Windheuser et al., 2023). Overfitting and bias are two other common deficiencies in ML algorithms. Overfitting is a major issue in ML and occurs when a model cannot generalize well on unseen data during the testing period. Over-fitted models tend to capture all the trends in the data, including unavoidable noise on the training set, instead of learning the hidden relationship between features. The main causes of overfitting are noisy data, insufficient training data, and model complexities. In this review paper, we discuss existing opportunities to address these issues and provide a path for future ML research in irrigation decision-making. This review paper aims to (i) assess how ML has been used as an intelligent simulation model to promote water management in the irrigation sector, (ii) discuss the results of ML models applied in various agricultural water use simulation research, and (iii) highlight the limitations that hinder the effective performance of ML models in irrigation prediction and scheduling. By doing so, we discussed the applicability of various ML algorithms in irrigation management, articulated their weakness and strengths, and identified the knowledge gaps and the next steps to promote the application of these intelligent algorithms in irrigation scheduling.

The remainder of this paper is structured as follows: Section 2 provides a brief introduction to ML as well as the most used algorithms in irrigation prediction. Section 3 presents surveyed studies that used ML approaches in irrigation-related studies. Section 4 presents existing limitations in the use of ML models and how they can be addressed, Section 5 highlights different ways to add physics into ML models, while Section 6 discusses major results from developed irrigation tools. Section 7 draws major conclusion of the paper.

2. Machine learning modelling concepts

ML, defined as "the study of computer algorithms that can improve automatically through experience and by using data" (Mitchell, 2007), has been used in agriculture due to the increased availability of data and the need to improve traditional farming practices. IoT is an emerging technology that uses smart sensors and devices interconnected through the internet to collect data and transmit them to a server with no or limited human intervention. As such, it allows continuous monitoring and management of soil and climate conditions and promotes efficient farm management. ML and IoT enable big data collection and modeling, and farm machinery modernization that gives rise to precision agriculture, also known as smart farming. Precision agriculture is the use of up-to-date information technologies, smart devices, and software tools to support decision-making in agriculture (Pierce and Nowak, 1999). It aims at increasing the farm's profitability by reducing the cost of production while improving environmental sustainability.

ML algorithms fall under three categories, as illustrated by Fig. 1: supervised learning, unsupervised learning, and Reinforcement Learning (RL) algorithms. In supervised learning, the developed model uses information from labeled inputs and outputs to find the functional relationships between them. Supervised learning can be used to predict categorical output values in which case a classification algorithm is used, on the other hand, if the value to be predicted is numerical, a regression algorithm can be used. In the unsupervised learning category, the learning algorithm uses unlabelled datasets to find the pattern that maps inputs to outputs. RL uses an agent to make a decision by interacting with the environment aiming at finding proper actions that maximize rewards and minimize penalties through trial and error. Deep learning (DL), a subset of ML, is also illustrated in Fig. 1. DL models can abstract features from high dimensional time series datasets and use multiple layers to progressively extract higher-level features from the

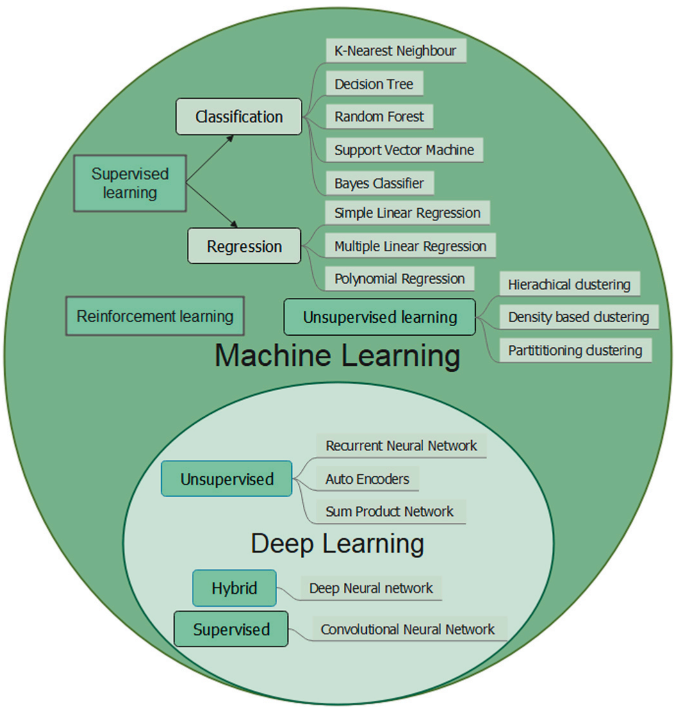

Fig. 1. Different types of ML algorithms.

raw input data. This review paper is mostly focused on broader ML models, even though some review examples of DL models for irrigation demand modeling are also provided.

Three key steps are followed in building ML models to perform prediction tasks (see Fig. 2). Firstly, the input data is pre-processed by carrying out data cleaning to remove noise and missing data using data transformation methods such as discretization or normalization. The model is then trained and calibrated with the training dataset to find the function that relates the inputs to the outputs by determining a set of parameters that give the best results (Mahesh, 2018; Oden Technologies, 2023). After this process, a new dataset (validation dataset) is used to evaluate the performance of the model. Once an overall satisfactory performance is achieved, the model can be tested on an unseen dataset, the performance of the testing period determines ML's ability to accurately perform a task based on the learned experience. Supervised learning and RL are recently used for irrigation water use modeling which are discussed in the following sections.

2.1. Machine learning algorithms commonly used in irrigation scheduling

Most ML algorithms for regression problems estimate a mapping function that can predict the output variables (Y) for new input data (X). Some popular examples of ML algorithms are Linear Regression, Random Forest (RF), Boosting Classifiers, Support Vector Machines, and Long Short-term Memory (LSTM). Here we briefly explain different ML models. Readers are referred to Kecman (2005), Kavitha et al. (2016), and Tabas and Samadi (2022) for more details on ML algorithms.

2.1.1. Random forest

Inspired by the work of Amit and Geman (1997), Breiman (2001) introduced RF. RF is a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1,\ldots\}$ formed based on a set of input vectors ($X = X_1, X_2, \ldots, X_n$) and random vectors ($\Theta_k$) sampled independently but with a similar distribution to find a prediction function ($f(X)$) that can accurately predict the variable of interest (Y). RF can be used to solve classification or regression problems. After multiple (k) runs, the majority vote of the predictions given by each tree is the answer for classification problems whereas the predictions' average is considered the model output for regression problems. Fig. 3 illustrates the workflow of RF simulation. The accuracy of the prediction function is given by the loss function $L(Y, f(X))$ which is a measure of the closeness of $f(X)$ to Y. Thus, the optimum prediction function minimizes the expected value of the loss function $E_{XY}(L, f(X))$.

2.1.2. Boosting classifiers

Boosting is a type of ensemble method that builds a good performing model (strong learner) by adding up several models in an attempt to combine multiple simulations until an accurate prediction is achieved or a maximum number of models are added. This is done in a sequential learning process where each model learns from the errors of its

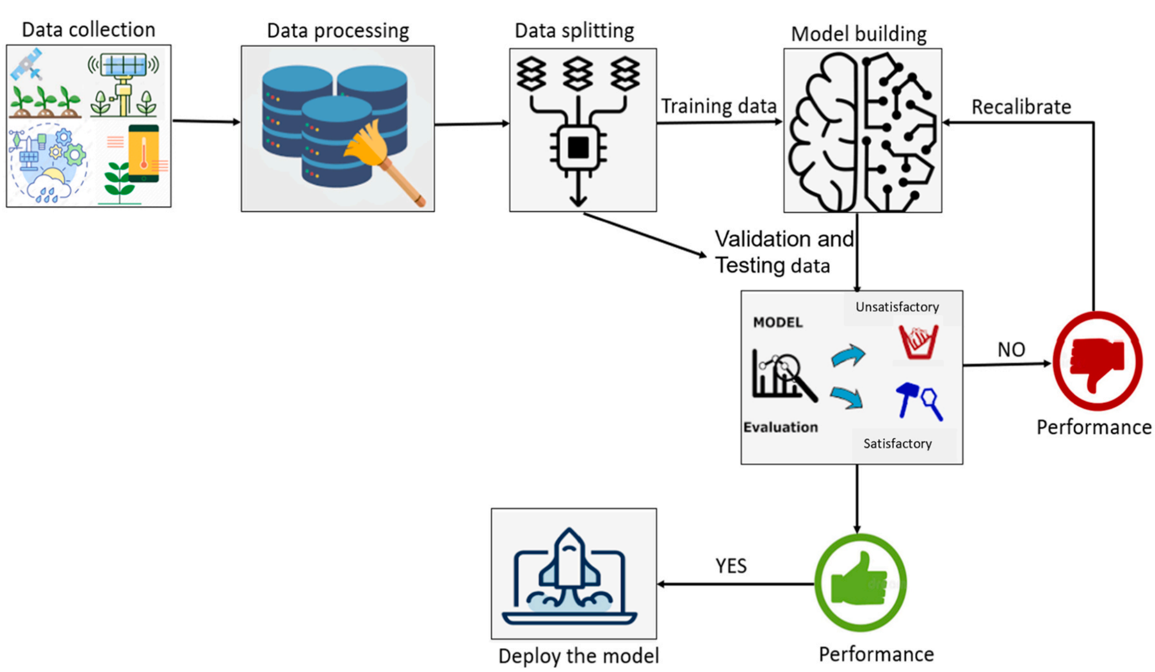

Fig. 2. The workflow of ML approaches for irrigation water use modelling.

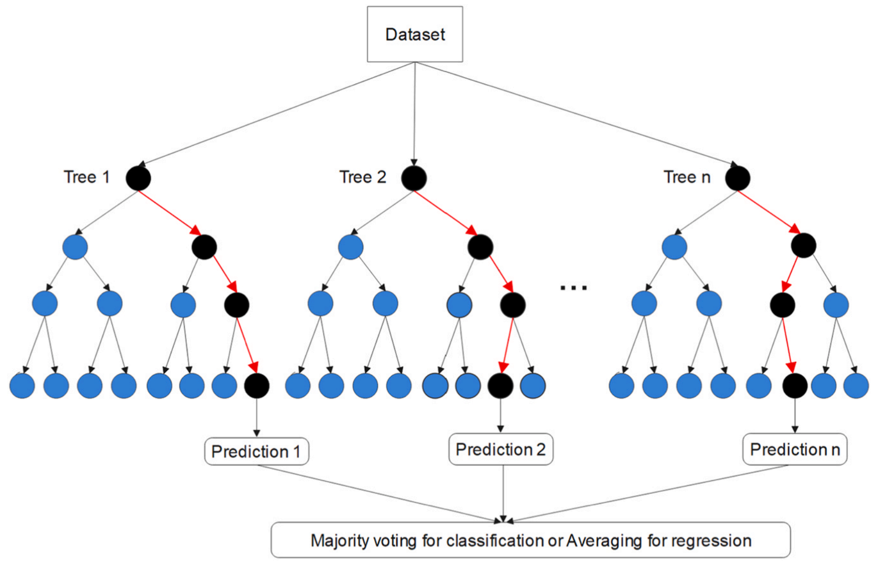

Fig. 3. The workflow of RF simulation.

predecessor and tries to correct them. Popular Boosting ensemble models include Adaptive boosting (Adaboost) and Gradient Tree Boosting (GTB). In the Adaboost, observations are weighted, and less weight is given to instances already well classified whereas those that are more difficult to classify are given more weight to be handled by new learners. GTB uses a loss function such as squared error to minimize the error as models are added up.

2.1.3. Long short-term memory

LSTM is a type of recurrent neural network (RNN) developed by Hochreiter and Schmidhuber (1997). Unlike the traditional RNN, LSTM networks allow information to persist, thus managing the problem of the vanishing gradient in the traditional RNN. The LSTM has a hidden state known as short-term memory and an additional cell state, which is the long-term memory. The main attribute of LSTM networks is that they can remember information for a long time. Fig. 4 illustrates the structure of an LSTM network. As illustrated, the information flow is regulated by three gates; the forget gate determines if the information from the previous time step should be kept or forgotten, the input gate learns new information from the current time step and adds it to the cell, and the output gate that transmits the updated information from the current time step to the next one.

2.1.4. Reinforcement learning algorithms

RL is an ML approach where the algorithm learns how to map situations to actions by interacting with its environment to attain a goal. This continuous interaction with the environment forms a closed loop

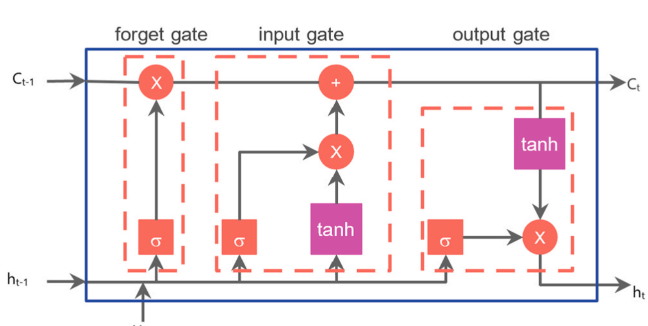

Fig. 4. The LSTM network structure. X_t is input, C_{t-1} is memory cell, and h_{t-1} is hidden layer. The tanh function is used to regulate the values in the cell state.

problem solving system where the selected actions depend on learned experience through a series of interactions with the environment and the rewards received per each action chosen (see Fig. 5). The goal of the RL algorithm is to maximize the cumulative rewards through the interactive feedback mechanism that differentiates RL from other ML algorithms.

RL uses a mathematical model, i.e., the Markov Decision Process (MDP), to solve a problem. MDP uses a property that enables the next state and reward to be computed given that the current state and action are known. More precisely, the agent interacts with the environment in a time sequence of discrete time steps. At each time step t, the agent receives the current state of the environment $S_t \in S$ with S being the set of all possible states and then selects an action $A_t \in A(S_t)$ where $A(S_t)$ represents actions that can be taken in state S_t. Following the action taken, the agent will then be given a reward $R_{t+1} \in R$ and the environment transits to a new state S_{t+1}. The goal of the agent is to maximize the cumulative rewards received from the environment. The measure of the values of possible action "a" at state "s" pairs that can be taken by the algorithm is determined by a value function $Q(s,a)$. This function can be optimized through a series of iterations to give the optimal state action pair that results in the highest reward at each time step. For complex environments with a high dimensional state-action space, the value function (Q-value) is not suitable for defining the reward of each pair. Consequently, a more advanced RL algorithm known as Deep Q-Network (DQN) consisting of a combination of Q-learning and Deep Neural Networks (DNN) is used to solve such complex problems. Instead of determining the cumulative reward based on the value function table, DQN leverages DNN to approximate the Q-values, the network takes the environment state as input and predicts the Q-values for each action.

Focusing on RL applications in irrigation water use simulation, researchers define the state as actual observations to determine a sequence of optimal irrigation that is related to a farm's environment such as soil moisture, crop growth status, and weather conditions. The agent, which is the learner and decision maker, is represented by the irrigation system while actions are the various choices made by the agent on the amount of irrigation water to apply. The reward function (R) is the profit made at the end of the season after the expenses associated with irrigating are covered and the goal of the agent is to perform a set of actions that maximizes this profit.

3. Literature review

The literature review focused on collecting a wide selection of refereed journals from major publishers. A total of 56 publications were

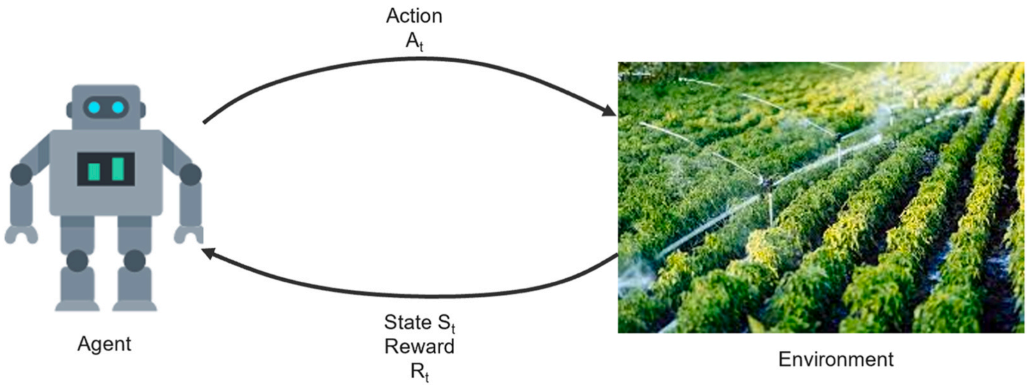

Fig. 5. RL agent and its interaction with a physical environment (i.e., a farm system).

initially found. Among them, 16 studies were selected as they were more relevant to the application of ML algorithms in irrigation prediction and decision-making. These 16 studies are discussed as follows.

Navarro-Hellín et al. (2016) were the first scholars to use Partial Least Squares Regression (PLSR) and Adaptive Fuzzy Inference Systems (ANFIS) to predict the weekly irrigation run time of three drip irrigated citrus plantations in Spain. The goal was to find the irrigation duration that maximizes the yield while optimally managing water. Using volumetric water content, soil water potential and crop evapotranspiration as input data, the predicted irrigation run time by the two models were evaluated against irrigation reports provided by an agronomist. Both models accurately predicted the irrigation pattern followed by the agronomist but ANFIS outperformed PLSR by giving a better estimation of weekly irrigation amount.

Sun et al. (2017) developed a RL algorithm to perform irrigation planning and scheduling. A neural network was used to emulate DSSAT in predicting seasonal total soil water content using irrigation and weather information as input. A second neural network was then used to predict crop yield using the predicted total soil water content. The crop yield prediction was then used as the training data to train the RL model. Soil moisture and yield were predicted to train the model and find the optimal irrigation schedule that maximizes the net return (see Equation 1). The net return was then taken into consideration in the model to measure economic gain and adapt to future changes in the produce and water price. The model was tested by simulating maize planted fields in Temple, Texas, (USA), Kunnunurra (Australia), Hyderabad, (India), and for wheat yield prediction in Saskatchewan (Canada). The predicted total soil water results were used by an irrigation controller to apply an adaptive amount of water as needed. This technique enabled automatic irrigation and surpassed threshold and fixed-based irrigation in the tested locations by achieving a higher net return. The developed irrigation system addresses the dynamic and complex nature of irrigation scheduling and control by taking into consideration the current and future soil water content variations to adaptively optimize irrigation schedules which is a significant contribution to improved water efficiency in irrigation systems. Its limitation, however, is that it is restricted to a small total soil water space and is not suitable for large-scale problems (continuous soil water space). Furthermore, the proposed system has not been tested with actual field conditions, therefore its performance cannot be validated in real applications.

Net return = yield (kg/ha) * price of product (dollars/kg) – water use (ha-mm/ha) * price of water (dollars/mm) Equation 1.

Kumar et al. (2017) created a fully automated intelligent irrigation system based on IoT to gather temperature, soil moisture, and raindrop sensor data. The data collected is used by a microcontroller for analysis of the irrigation needs of several crops including beans, curry leaves, lady fingers, moringa, and radish. A motor is activated for irrigating the field when moisture values drop below a threshold range. Using recorded temperature and water flow data, a linear regressor is used to predict the next irrigation water requirement. The system then schedules irrigation based on the irrigation amount predicted by activating/deactivating a relay which in turn activates or deactivates a solenoid valve as needed. The linear regression model is updated in case there is a change in the data. The system also includes a mobile app that enables users to have access to field data remotely. The performance of the system was, however, not compared to other irrigation demand assessment approaches, nor did the authors specify how much water can be saved. This work contributed to advancing the automation of irrigation systems to enable remote monitoring and control. Its innovative incorporation of IoT technology with ML for water demand assessment and scheduling was a stepping stone for achieving better results in water use optimization with limited labor.

Goap et al. (2018) developed a smart irrigation system based on IoT and ML techniques. A combination of sensor-collected field data such as soil moisture, soil temperature, Ultraviolet (UV) light radiation, and weather forecast data including precipitation, air temperature, and other variables, was used to compute future soil moisture based on which irrigation requirements were determined. The smart irrigation architecture consists of an IoT data collection system that collects, transmits, and processes sensor data and weather forecasts. A soil moisture prediction algorithm was developed as a combination of SVR and k-means clustering algorithms that calculate soil moisture change due to weather conditions. An irrigation scheduling algorithm then takes the predicted daily soil moisture, retrieves precipitation information to find the nearest day it is expected to rain, and computes the required soil moisture to maintain crop growth. If the required soil moisture is below the minimum threshold a command is sent to a relay switch to start the water motor and vice versa. A responsive web portal is used to view the projected soil moisture along with expected rainfall and to start or stop irrigation. The algorithm achieved a correlation coefficient and Mean Squared Error (MSE) of 0.56 and 0.135 respectively, when estimated soil moisture deficit (SMD) was compared to sensor-based SMD. The limitation of this work is that no water-saving analysis was performed to assess its potential against other systems. Additionally, the experiment was performed in a garden, so there is no record of its performance on a larger area such as a farm system. On the other hand, the study exemplifies how fully autonomous smart irrigation systems can be developed using sensor-collected data and weather forecasts to run ML algorithms and predict water demand.

Concurrently, Goldstein et al. (2018) used regression and classification algorithms to build an ML model based on sensor and weather data to predict irrigation quantities for jojoba-planted fields in Israel as prescribed by an agronomist. These data were subdivided into eight subsets of different features to determine the accuracy of the models when inputs were changed. LR, Gradient Boosted Regression Tree (GBRT), and Boosted Tree Classifier (BTC) were used to estimate irrigation water demands. The models were evaluated on their ability to provide an irrigation value that is not more than 0.2 mm above or below the agronomist's one. It was observed that LR performed unsatisfactorily giving a Root Mean Square Error (RMSE) equal to 0.466 and a performance accuracy of 52.3%. GBRT outperformed LR as the lowest RMSE was 0.11 and the accuracy rate reached 92.7%. The BTC algorithm also performed well with an accuracy of 95%. It was concluded that non-parametric models perform well compared to LR methods in cases where there is a non-linear functional relationship between the dependent and independent variables. The authors also observed that the model run with less input data (such as the model run without saturation/drought data) provided quite similar results to the model in which all the variables were used to train the model highlighting the fact that selecting appropriate features is more important in achieving satisfactory performance.

In another study, Adeyemi et al. (2018) developed a predictive irrigation scheduling system using the Feed Forward Neural Network (FFNN) model and LSTM. A neural network (NN) model was used to predict the one-day-ahead volumetric soil moisture using historical soil moisture and climatic data. Based on the predicted soil moisture, crop water requirement, and soil water retention, the irrigation depth and timing were determined for a potato field. According to the obtained results, the two models performed comparably on the validation datasets and provided better predictions than those of a non-ML approach that estimates the soil moisture from the average of previously observed three-day soil moisture values. However, the LSTM model outperformed the FFNN model in terms of regression coefficient when datasets from different sites were used. This showed the robustness of LSTM networks in approximating the underlying functional relationship between inputs and outputs for situations similar to the learned ones. The LSTM model was coupled with AQUACROP to simulate potato crop growth and develop a predictive irrigation scheduling system whose goal was to maintain the soil moisture within a set of upper and lower boundaries during the simulation period. Given the predicted soil moisture, the deficit amount of water to reach the upper bound was computed and the irrigation amount was defined as the water depth needed to replenish the soil to attain the set threshold. The system was compared with a rule-based irrigation system, the former achieved a water saving of more than 20% in the three tested sites compared to the latter while maintaining the same water use efficiency.

Using daily climatic data and daily irrigation amount collected in a period of one year, Perea et al., 2019 developed a model to predict irrigation amount based on farmers' decisions about when to irrigate. The crops grown in the area were drip-irrigated tomato and maize and rice, which was flood-irrigated. The purpose of the model was to simulate the occurrence of irrigation one day ahead as planned by farmers. This model was developed using DT and optimized using a Genetic Algorithm (GA) to find the optimal DT that accurately simulates farmers' behavior. Based on the defined objective functions, three Classification Trees (CT; CT1, CT2, and CT3) were selected. The first CT (CT1) outperformed other DTs by accurately predicting all the observed irrigation and no irrigation events, achieving a performance accuracy of 100% in both cases. CT2 followed with an accurate classification of 73% of the irrigation rates and 93% of no irrigation rates. Lastly, CT3 was able to classify 68% of the actual irrigation rates and 93% of no irrigation rates. The developed model contributes to irrigation management by providing information on when to schedule irrigation and hence could help in better coordinating irrigation activities to minimize energy and water loss.

Weekly irrigation amounts of orchards of citrus trees (orange, mandarina and lemon trees) located in Southeast Spain were predicted using a modelling system driven by LR, a random forest regressor (RFR), and SVR algorithms (Torres-Sanchez et al., 2020). These algorithms were trained to perform the irrigation assessment, as such the goal was to imitate the irrigation recommendations of an agronomist. Using the daily average matric potential of the previous five days, the total water needs, whether the fruit is gaining weight or not expressed as a binary value as input the ML models were trained to output the total irrigation amount recommended by the agronomists for the following week. The results were compared with irrigation prescriptions by agronomists. Results showed that RFR performs better than SVR and LR with an RMSE of 16.83m3, 17.13m3, and 19.5m3, respectively. A weekly average error of 9% was obtained when the predicted irrigation amount was compared with the agronomist's report.

To overcome the drawbacks of traditional RL models such as Q - Network (QN), Yang et al. (2020) proposed a deep Q - Network (DQN) to control irrigation scheduling. The difference between QN and DQN is that the former generates the value function table while learning from the data whereas the latter uses ANNs to compute the value function from multi-dimensional input data. Hence ANNs offer the possibility to handle a large amount of data and precisely render the irrigation scheduling problem more scalable. The environment was interfaced by AQUACROP, and simulations were conducted using three crops namely maize, wheat, and soybean. The state of the environment was described by the date, crop stage, precipitation, reference ET, total water content in the effective root zone, stomatal water content, and irrigation. Actions consisted of irrigation amounts from 0 mm to 10 mm with a 0.5 mm increment. Results demonstrated that the suggested algorithm outperformed traditional irrigation approaches such as constant irrigation, and percentage available water in dry, moderate, and wet weather conditions and was closest to the Q learning estimation for dry conditions. Coupling AQUACROP with DQN for irrigation scheduling allowed for a more comprehensive assessment of crop water demand by enabling the simulation of different crop growth scenarios based on the applied irrigation amount. This integration also addresses the data scarcity issue often encountered in ML modeling. The system was however not directly tested on the farm, the replication of results as far as implementation is concerned is thus not ascertained.

To determine the right set of features to use for determining the irrigation scheduling of a drip-irrigated rice field from weather parameters based on ML, Sidhu et al. (2020) used correlation to identify the dependencies in the data and guide the selection of feature sets. Features with a correlation of 0.7 and above were discarded. The variables found useful and used as input features for irrigation scheduling were number of days after sowing, maximum temperature, dry bulb at 6:00 AM, humidity at 2:00 PM, soil temperature 20 cm below surface at 2:00 PM, sunshine hours, wind speed, wind direction, ET, number of days since last rain and number of days since last irrigation. Among different algorithms used in their study, Decision Tree Regressor (DTR) and Adaboost outperformed other algorithms. As noted by Goldstein et al. (2018), LR was unable to perform well in predicting irrigation demand. It was also demonstrated that Adaboost could perform equally well with a limited number of selected input features.

Zhou (2020) proposed an RL method that uses a Convolutional Neural Network (CNN) to estimate the value function used to train the DQN algorithm. The algorithm uses sensor-collected soil data such as moisture and humidity, air temperature and air humidity, leaf water potential, and leaf conductance collected from a grape-growing greenhouse plantation as input features to train the model. The system improved irrigation decision-making on the plantation with no human intervention.

To overcome the limited learning capability of some previously developed ML models, Kashyap et al. (2021) created a Deep Learning Intelligent Irrigation System for precision Agriculture (DLiSA) based on NN and IoT-enabled approaches that account for climate and soil moisture variations. The system comprises the one-day-ahead soil water content predicted by an LSTM model, an irrigation scheduler to compute the volume of irrigation water, and an irrigation planner to calculate the duration of the irrigation period and the water delivery across the field. Irrigation commands are then passed to the actuator nodes from which they get to the water valves. The particularity of this system is that it integrates feedback from climate and soil sensors which enables it to adapt to local changes and perform well during historical periods. A comparison of this irrigation system with those obtained with a threshold-based irrigation model and an FFNN showed that 43% and 23% of water volume, respectively are saved using DLiSA.

Jimenez et al. (2021) used LSTM with soil matric potential (SMP) collected at three soil depths (i.e., 150, 300, and 600 mm), rainfall, and irrigation amounts as input data to predict irrigation schedules during a corn-growing season for sandy clay loam and loamy fine sand soils for 1, 2, 3, 6, 12 and 24 h in advance. In addition, three different LSTM models were developed by changing inputs to evaluate the predictability of LSTM networks concerning the input data. The input data used for the first model were SMP at 150-, 300-, and 600-mm soil depths, management zone, rain, and previously applied irrigation; the second model used SMP and management zone, whereas the third model only used SMP as input. The results indicated that the same error and predictability level obtained for the first model can also be obtained using only SMP data. In terms of temporal models' performance, the 1-hour ahead irrigation predictions were accurately predicted the same as the three and six hours-ahead, but the accuracy gradually decreased for irrigation predictions of 12- and 24-hours advance causing overfitting or underfitting. Nevertheless, it was concluded that irrigation prescriptions changes can be accurately forecasted on time by LSTM. The authors also noted that using rainfall and irrigation data as input caused overfitting in the testing period. This is because rainfall patterns change over time and those rainfall amount fluctuations might be unknown for the model.

Y. Chen et al., 2021 combined several classifiers to form a strong irrigation predicting approach using stacking and boosting ensemble learning methods. The model was developed to predict irrigation volumes for multiple fields in a greenhouse of organic vegetables. Air temperature, air humidity, soil temperature, soil humidity, and light intensity data were used as input variables to linear SVR, Support Vector Classification (SVC), RF, and DT. Model performances of 0.45, 0.675, 0.645, and 0.54 were obtained for each model, respectively. Stacked generalization was employed to combine the advantages of these algorithms using extreme gradient boosting (XGBoost) which improved the performance from 0.58 to 0.64. The best performance was obtained when simulated and observed water volumes were compared with 3.33 and 8.16 for the MAE and RMSE, respectively.

Concurrently, Deep Q learning was implemented by M. Chen et al., 2021 to perform irrigation scheduling for paddy rice using short-term weather forecasts. The input features defining the field environment on a daily basis consisted of one week forecasted rain sequence, water depth as well as minimum and maximum thresholds of water depth and the maximum allowable water depth where the maximum allowable water depth and the maximum threshold are determined based on the growth stage. The three irrigation actions that are defined to irrigate were no irrigation, supplying 50% or 100% of the water needed to attain the maximum water depth. Results showed that no yield losses were experienced in the studied period. The DQN was able to consider current field water conditions and weather forecasts to efficiently choose to irrigate or delay irrigation in order to avoid under or over irrigation which resulted in water conservation and increased rainfall use. Comparing irrigation with the flooded irrigation strategy, the DQN method reduced water use by 23 mm. This research demonstrates how advanced computational techniques based on ML can enhance irrigation prediction. By applying weather forecasts in crop water demand assessment to limit unnecessary irrigation, more water is conserved compared to conventional irrigation methods. The study also underscores the ability of the applied algorithm to optimize water despite imperfect weather forecasts thanks to its ability to learn from past irrigation experiences and weather forecast uncertainties.

More recently, Alibabaei et al. (2022) trained a DQN to schedule and optimize irrigation for a tomato field. The state of the environment was expressed by a combination of average temperature, average relative humidity, precipitation, reference evapotranspiration, average wind speed, total soil water in profile, and irrigation amount. Actions consisted of irrigation amounts ranging from 0 to 60 mm - ha/ha. The agent environment was simulated by two LSTM models. One was to predict the one-day soil moisture in advance and the other was used to approximate the seasonal harvest based on the growing season environmental conditions and then determine the net return. To estimate the Q-table, an ANN, an LSTM, and a CNN were used. LSTM surpassed other algorithms because it memorized previous input data and used the retained information in the subsequent time steps. The results achieved by the developed algorithm showed an increase in the productivity in the test field and a reduction of irrigation water in comparison to threshold-based and fixed irrigation by 18% to 30%. The crucial finding of this study that showcases the strength of ML models is the ability of the developed irrigation system to adapt irrigation scheduling to the crop growth stages. At the beginning of the season, the system avoids water losses by adopting a lower water supply scheduling which gradually increases as the season progresses and eventually reduces to stress the plants towards the harvest period. This ensures that enough water is available to sustain the plants during the critical growth period and minimizes water losses at the start and end of the season.

The surveyed studies illustrate the ability of ML algorithms to make use of data from different sources including weather forecasts, to assess crop water demand and predict irrigation schedules in a more efficient way that optimizes water use compared to conventional approaches. The attained water efficiency is shown by the amount of water saved as opposed to traditional fixed period or threshold value irrigation techniques that do not account for spatial-temporal changes in soil, crop, and meteorological variables. With the real-time data acquisition approaches enabled by IoT and the progress made in coupling IoT with ML models as applied and demonstrated in the several presented studies, ML algorithms hold great potential for automating the irrigation decision-making process. This will improve further timely irrigation scheduling and eradicate the additional work of manually controlling irrigation systems. Applying ML models in irrigation decision-making and scheduling is also economically beneficial as the crop water waste, the cost of water, and labor demand are minimized. Table 1 summarizes the reviewed studies, the algorithm used, and the obtained model performance.

4. Current limitations of machine learning application in irrigation water use modeling

4.1. Data scarcity

The performance of ML models highly depends on the availability of sufficient data to accurately learn the underlying relationship between the predictors and the variable to be predicted. Due to the high variability of soil and climatic conditions in space and time, data should be available at a fine spatial-temporal resolution for proper results to be obtained. Data availability is often a challenge as the collected data by sensors such as soil moisture, are point measurements and hence do not represent the entire field's conditions. To obtain representative information about the field's conditions and improve the performance of ML models, many sensors need to be installed across the field which can be expensive. In addition, satellite-derived data can be useful for irrigation prediction but often are not available at the field scale. Satellite data can be an alternative source of data, however, the frequency at which data is collected is usually low due to orbital paths. The collection of good-quality satellite data is also hindered by the effect of canopy or cloudy atmospheric conditions which leads to significant data gaps. Moreover, the spatial resolution of these data is sometimes not satisfactory. For instance, the presently operating soil moisture products are available at a coarse resolution that cannot be directly used in irrigation prediction at the field level if not downscaled (Zhang et al., 2021). On the other hand, downscaling data adds more uncertainties that can compromise the modeling results.

The data scarcity problem can be addressed by integrating remote sensing and field sensors' collected data to increase the reliability of simulated results. In the event of low sensor coverage, satellite-based

Table 1
ML based irrigation studies reviewed in this paper.

Reference	Aim	Input data	Functionality	Crop	Model (s)	Performance
Sun et al. (2017)	RL-based irrigation control system for the optimization of net return	Soil water content, ET, rainfall TSW, irrigation Predicted crop yield	Total soil water content crop yield Irrigation planning and scheduling	Maize and wheat	1st NN 2nd NN QN	ML method outperformed threshold based and fixed irrigation by 46.7% and 59.8% on average net return, respectively
Kumar et al. (2017)	Automatic monitoring and control of irrigation	Temperature, soil moisture, and rainfall	Irrigation requirement forecasting	beans, curry leaves, ladies finger, moringa and radish	LR	Not mentioned
Perea et al. (2019)	Emulate farmers' irrigation decisions	Climate data and irrigation amount	Simulation of irrigation occurrence 1 day ahead	Tomato, maize and rice	DTs	68-100% irrigation events are positively predicted.
Kashyap et al. (2021)	ML models in unpredictable climates	Soil moisture and climate data	1 day ahead soil moisture, variation of water application spatially, irrigation period	Grassland, farmland and arable	LSTM	Saved up to 43% volume of water compared to threshold-based irrigation.
Adeyemi et al. (2018)	soil moisture at 3 sites of different soil types	Soil moisture, climate data, rain data, irrigation	irrigation amount prediction using 1 day ahead soil moisture forecast	potatoes	LSTM	46, 20, 31% of water saved at each site
Goap et al. (2018)	Possible soil moisture values prediction to aid in irrigation planning	soil moisture and temperature, UV light radiation and weather forecast	Computation of soil moisture changes and irrigation scheduling	garden	SVR and *k*-means	$R^2 = 0.56$ MSE = 0.135
Goldstein et al. (2018)	Weekly irrigation scheduling forecast as determined by an agronomist	Soil moisture and weather data	Irrigation recommendation	jojoba	LR, GBRT and BTC	Success rate- GRBT= 93% Success rate- BTC= 95% Success rate- LR= 52.3%
Torres-Sanchez et al. (2020)	Predictions capabilities of several learning models with experts' decisions	Daily average of SMP, total water needs, applied water 1 week before, crop growth stage	1-week irrigation amount prediction of orchards	Citrus (Orange, mandarin and lemon trees)	RFR, SVR and LR	RMSE-RFR= 16.83 m^3, RMSE-SVR= 17.13 m^3, RMSE-LR= 19.5 m^3
Sidhu et al. (2020)	Traditional ML methods for irrigation scheduling	Weather and soil variables	Predict rice water demand	rice	LR, SVR, RFR, DTR, Adaboost, GBR and NN	Best performance was obtained with Adaboost; MSE= 125.79, $R^2 = 0.79$ Accuracy= 71%
Yang et al. (2020)	Irrigation scheduling approach with deep learning	AquaCrop simulations, and weather data	DRL irrigation scheduling model	Maize, wheat and soybean	QN, DQN	Highest net return obtained for corn and wheat; second highest return obtained for soybean
Jimenez et al. (2021)	Irrigation prescriptions requirement	SMP, irrigation management zone, rainfall, and irrigation	irrigation amount prediction for 1, 3, 6, 12 and 24 h ahead	corn	LSTM	$R^2 = 0.82$ - 0.98 for 1 h ahead prediction
Y. Chen et al., 2021	Irrigation volumes prediction	Air temperature, and humidity, soil temperature and humidity, light intensity	Irrigation volumes	vegetables	SVR, SVC, RF and DT	SVR= 0.45 SVC= 0.675 RF= 0.645 DT= 0.54
M. Chen et al., 2021	Irrigation decision making based on weather forecast	Rainfall forecast and water depth in the field	Irrigation scheduling	early, middle and late rice	DQN	Reduced irrigation water by 23 mm on average and reduced drainage water by 21 mm
Alibabaei et al. (2022)	Irrigation optimization	Climate data, irrigation amount, total soil water in profile	Irrigation scheduling	tomato	DQN	Decreased irrigation amount by 20 – 30%

measurements can help bridge the spatial-temporal data gap. Satellite data such as soil moisture, surface temperature, Leaf Area Index (LAI), etc. can be used to supplement sensor data in modeling field conditions after carefully downscaling the data to the field level. Satellite-based measurements can also be used to validate sensor-measured data by detecting outliers. Having a complete, reliable dataset would considerably improve the performance of ML models and allow farmers and Extension irrigation specialists to make more informed irrigation decisions.

Furthermore, most models use soil moisture-based irrigation scheduling methods; thus, soil moisture is commonly used as an indicator of crop water stress. However, plant water stress-related indices can also be useful predictors to assess irrigation demand (Jones, 2004; Gu et al., 2020) but no significant work has been done to build ML models for irrigation predictions using these variables. Thus, data of either plant-water status, such as measurements of leaf/stem/xylem water potential and leaf thickness, or plant physiology, such as sap flow measurements by sensors, stomatal conductance, and plant thermal sensing, can potentially enhance data availability and irrigation prediction accuracy.

In addition, data augmentation techniques have been proven to remediate the data scarcity problem. Specifically, image data can boost the performance of ML models in irrigation modeling and applications. Data augmentation techniques are mostly applied in crop-weed classification (Fawakherji et al., 2020; Su et al., 2021; Divyanth et al., 2022; Espejo-Garcia et al., 2023) and plant disease diagnosis (Wu et al., 2020; Cap et al., 2022; Huang et al., 2022) Data augmentation involves implementing a sequence of methodologies to generate new images from the existing image dataset. Data augmentation techniques fall into three categories; the first class involves basic augmentation techniques that increase the size of a dataset by changing the geometric position of points in an image hence generating a new data point examples of this technique are geometric transformations such as scaling, rotating, and skewing images. Noise injection techniques such as Gaussian or uniform noise are also common techniques to increase data samples. They entail adding noise to data to increase the variance. The deformable augmentation technique is the second category of data augmentation techniques that can be used when the basic techniques do not provide satisfactory variability and include methods such as spline interpolation, which uses a piecewise polynomial function to interpolate a new data point between two observed points. Using statistical shape models is another advanced method to generate new data by using observed data and a mathematical model to understand the dataset's variability and implement changes within the limit of possible parameters. The third category of data augmentation techniques is generative DL techniques. These approaches increase the volume of data by learning features in a dataset to generate synthetic but realistic datasets. The most common DL networks are generative adversarial networks (GAN) and their variants. Adversarial learning refers to the contest between two NNs, a generator and a discriminator. The generator learns to generate new data such that the distribution of the generated data is as close as possible to that of the observed data. The discriminator, on the other hand, learns to evaluate the generated data and differentiate them from authentic data. In addition to increasing the sample size, these techniques reduce overfitting by introducing new patterns in the data that can make a model more generalizable.

4.2. Noisy and heterogeneous data

Data is the foundation of every ML project. Hence, preprocessing data prior to their use is an important step to ensure data quality and reduce the possibility of defective data that hinder model efficiency. Data noise and heterogeneity are the common issues encountered during data collection and preprocessing. Raw data contains numerous flaws such as missing values, incorrect values, and incorrectly formatted values that introduce noise to ML. Directly feeding such data to ML models is inappropriate as they may fail to derive the right pattern. Poor data quality can thus lead to significantly unsatisfactory performance. Data should be machine-readable, trustworthy, and impartial. By ensuring data is correctly classified, data curation can assist ML model developers in validating the diversity of training data for irrigation modeling applications. Therefore, it is important to statistically analyze and understand the data and select suitable methods for removing noise and bias in the data. To counteract the substantial bias that data preprocessing methods may induce in predictions, it is recommended that maximum entropy constraint be applied on the inference step (Pfeiffer et al., 2015), forcing predictions to have a distribution similar to the one of observed values. Although data sparseness might persist and perhaps be exacerbated by big data, the volume of big data presents exceptional prospects for predictive analytics because sufficient frequency may have been achieved for diverse sub-samples. On-farm big data are in divergent formats, from distinct population samples, and so are very heterogeneous. It can be very tedious to transform heterogeneous data into a uniform format. In addition, these heterogeneous data may be of different levels of importance in the learning process (Zhou et al., 2017). Therefore, adding all the features and giving them equal weights in ML models is unlikely to lead to optimal learning results.

4.3. Lack of accessibility and reproducibility of data and machine learning models

Irrigation data is part of large earth system datasets that are produced at the farm scale; however, the majority of this data is not openly shared and is unavailable. Other irrigation metadata are put into password-protected and key-value data structures that limit their findability and accessibility significantly. Open access to data should be improved to better serve irrigation scheduling applications (Gu et al., 2020). Open data advances research by preventing duplicate data collection efforts freeing up resources to gather a more distinct collection of data and a more comprehensive record of observations. Secondly, open data regulations promote increased data usage and reuse. For example, after enacting the Landsat open and free data policy, data downloads increased by 20 times from 2009 to 2017, and, according to the annual number of publications, data use increased by four times (Z. Zhu et al., 2019).

Advancing research is not solely limited to making data readily available but also requires the availability of developed software. Providing open access to software's source code allows computational and scientific transparency and reproducibility. Scientific reproducibility is evasive if free access to code is restricted. Gil et al. (2016) argued that simply providing textual explanations of algorithms and code implementation may hamper reproducibility. Moreover, open-source software favors software reuse which reduces duplication possibilities, promotes the use of open data, and guarantees the code's lifespan (Committee on Best Practices for a Future Open Code Policy for NASA Space Science, 2018).

A rising recognition of the necessity of making data more findable, accessible, interoperable, and reusable (FAIR; Wilkinson et al., 2016) has set a standard for expanding the availability of research standards and tools that ease data sharing and management (Wagner et al., 2022). FAIR principles imply that data and tools must be easily findable, freely obtainable, and reusable so that methodology and derived results can be entirely transparent (Koymans et al., 2020). Data and tools can be findable when they are sufficiently described by rich metadata and registered or indexed in a known and accessible searchable resource. Accessibility means that users and machines, following appropriate authorization and via a well-defined protocol, can obtain data and the related tools. For data and tools to be interoperable, they should be described using normative and community-known specifications. Reproducibility is another aspect of the FAIR principle that received less attention in the irrigation community. Reproducibility necessitates describing in detail the characteristics of the data and tools, and their origin according to community standards with clear and accessible conditions for usage (Boeckhout et al., 2018). At a minimum, ML modeling outcomes should be reproduced and ideally replicated by other users before it is deployed.

Currently, farm-related data, including irrigation, soil moisture, and climate information, are not based on the FAIR principles, nor their metadata format and terminology are constrained by these guidelines. To move towards a FAIRer future that boosts the availability of farm-related data and tools, continuous generation of FAIR metadata standards and reproducible tools are indeed needed for future data-driven applications in irrigation modeling. Towards this end, scientific communities can adopt several best practices, namely making their data and ML algorithms accessible in a public cyberinfrastructure and in a non-proprietary standardized format, creating a digital object identifier for their data and tools, and providing clear licensing information along with constraints on the use. To encourage reusability, software and code should also be openly available through a non-restrictive license adopted by the community to facilitate reproducibility.

4.4. Blackboxness of machine learning models

The explainability and interpretability of ML models are another aspect of ML that has been questioned in the past (Kaur et al., 2020; Moraffah et al., 2020). Enhancing models' predictability comes at the cost of increasing their complexity which often makes them less interpretable. However, the interpretability of ML models is more important than their accuracy (Rudin, 2019). For example, to trust the irrigation recommendations provided by a model, irrigation specialists and crop consultants would be more interested in knowing the accuracy of irrigation recommendations for a certain scheduling time than in the structure of the algorithms and the types of performance metrics used in the model. Explainable ML has initiated a set of methods to increase the interpretability and explainability of ML models (Carvalho et al., 2019; Miller, 2019). In the agricultural domain, these techniques have been applied in several prediction tasks such as crop yield estimation (Sihi et al., 2022; Wolanin et al., 2020), prediction of daily pan evaporation and evapotranspiration (El Bilali et al., 2023; Zhao et al., 2023) and in monitoring fields' conditions for optimizing crop production (Sabrina et al., 2022). To interpret the results of a model, the current ML approaches aim to evaluate the effects of various combinations of input variables on the results by applying techniques such as Shapely Additive exPlanations (SHAP) that can assess the pros and cons of an input feature on the model output. Local Interpretable Model-agnostic Explanations (LIME) is another popular technique applied to make ML models' results human interpretable by estimating the behaviour of a complex model, it consists of adding noise to a data point without changing the rest of the dataset, fitting the new dataset to the model, and observing the behaviour of the model (Ribeiro et al., 2016). To the best of our knowledge, these approaches have not yet been adopted in explaining irrigation predicting models; there is, therefore, room to improve the understanding of these models by incorporating such techniques into the current ML modelling structures.

4.5. Lack of human interaction with machine learing-based irrigation modeling

Although considerable effort has been made to develop irrigation decision-making models, current models lack a way to integrate expert knowledge into the simulation process. The limited success of these models is partly because the decision rules used in the models are limited in scope than the factors growers use to make decisions (Car, 2018). Researchers mostly create models that can tackle agricultural problems without the need for human interaction. Unfortunately, a majority of existing models have not yet reached this degree of intelligence. Due to calculation time constraints and the complexity of agricultural problems, models may provide imprecise irrigation recommendations. As a result, agricultural knowledge from experienced irrigation specialists, Extension professionals, and experienced farmers can be employed in the model to improve the feasibility of ML outcomes and validate the results. An interactive interface that allows professionals to share their knowledge and perspectives can be an approach to considering expert judgment.

Another way to allow human control over ML models is to develop systems that enable human-in-the-loop feedback and interactions. The effective application of ML models in irrigation scheduling systems requires that humans control their outcomes to minimize the risk of lethal results. Efficient human control is the capability to incorporate human knowledge into the modeling process to make informed, timely decisions that provide the greatest operational outcomes possible (Boardman and Butcher, 2019). ML systems must create models that take into consideration a range of contextual data in order to function well in complicated settings (National Academies of Sciences, 2021). This allows the systems to make decisions by accurately understanding the current situation and projecting future scenarios. However, human-in-the-loop techniques need to be developed to ease the collaboration between ML models and irrigation experts such as Extension irrigation specialists and experienced farmers. This will help align goals and synchronize task prioritization, resource allocations, and improved irrigation decisions.

4.6. Uncertainty and error in machine learning simulation

Although ML techniques for irrigation decision-making were proven to be helpful in irrigation planning, all the reviewed papers did not take into consideration the errors and uncertainties in the predictions. Uncertainty in irrigation modeling stems from the continuously changing environment, such as meteorological conditions and soil moisture variations. The lack of uncertainty quantification in the ML predictions may result in overfitting or underfitting outcomes. Considering ML predictions without assessing the extent to which results are credible given the associated uncertainties in the input data and model often leads to inappropriate or biased decisions. For example, if the input data fed to a model trained to predict the irrigation amount lies outside the training data distribution, the model may generate unreasonable irrigation recommendations that are essentially based on an "*educated guess*". But, if the prediction is accompanied by confidence intervals that is the certainty the model itself has of its prediction, the decision taker would be informed of this randomness and decide on the irrigation amount accordingly. Thus, an assessment of uncertainties in the simulated results is crucial to ensure that the results fall within a certain level of confidence.

Generally, uncertainty can arise from input data and the ML procedures or structure in the modeling process. There are two types of uncertainties, (i) epistemic uncertainty resulting from the ML architecture choice and model parameters and (ii) aleatory uncertainty due to a physical system's intrinsic randomness. Aleatoric uncertainty describes the inherent randomness of the data-generating process, which cannot be justified by gathering additional data samples or observations (Kendall and Gal, 2017). This form of uncertainty is caused by unknown processes that pertain to the concept of randomness, i.e., the unpredictability in modeling outcomes. Aleatoric uncertainty can further be subdivided into homoscedastic and heteroscedastic uncertainties. Homoscedastic uncertainty is independent of ML inputs, which means that it remains consistent regardless of input. Because this sort of uncertainty produces uniform variances for all inputs, the variation in observation noise for all input parameters is constant (Tabas and Samadi, 2022). On the other hand, heteroscedastic uncertainty varies with datasets, where some inputs provide more noisy outputs than others.

Aleatoric uncertainty does not increase for out-of-distribution data contrary to epistemic uncertainty that can be handled in the modeling process if sufficient training data is available. Epistemic uncertainty may increase during the testing period for those data points that lie outside the training dataset. Thus, the larger the training dataset the lower the epistemic uncertainty. Modeling these two types of uncertainties increases the predictive performance of models. Understanding how individual sources of uncertainty propagate to the integrated irrigation system response is useful in analyzing the impact of data and model uncertainty on the simulated results. Additionally, distinguishing these uncertainties helps understand the sources of errors that have the potential to be reduced. Differentiating them also makes decision-making transparent and helps in making more informed (irrigation) decisions as unreduced uncertainties are clear and known (Der Kiureghian and Ditlevsen, 2009).

Uncertainty estimation methods can be classified into four categories namely single deterministic methods, Bayesian approximations, ensemble learning, and test-time data augmentation methods. The most widely used methods are Bayesian approximations such as Monte Carlo dropout, Markov Chain Monte-Carlo (MCMC) optimization methods (Duane et al., 1987), Bayesian Model Averaging (Samadi et al., 2020), and ensemble learning methods such as deep ensemble, Bayesian deep ensemble, and Dirichlet deep networks (Abdar et al., 2021). Largely, data uncertainty for a normally distributed uncertainty can be measured by predicting the parameters of a probability distribution for instance the mean and standard deviation (Kendall and Gal, 2017; Lakshminarayanan et al., 2017). These approaches can be coupled with ML-based irrigation water use modeling to reduce errors and uncertainty during both training and testing periods. However, to efficiently reduce uncertainty in ML modeling, rigorous tests of the suitability of different irrigation process parameterizations and model parameters are needed across different model development stages as well as various testbeds and farm settings.

5. Physics-guided machine learning (PGML)

PGML offers a compelling solution to improve irrigation decisions and the interpretability of ML models. The interactions between plants, the soil, and the environment are complex and require a wide range of physical processes to be accurately programmed and described in space and time by a data-driven model. PGML provides an opportunity to leverage the best attributes of both data and physical processes in the ML models (Shen et al., 2018). Combining physical processes with data-driven capabilities can be performed in a variety of ways. Approaches to incorporate physical principles into ML models involve constructing new ML architectures that can employ some physical laws and processes in the irrigation water use prediction. These improvements are possible because ML models are modular and versatile. Scientific knowledge can be utilized to define node links that capture physics-based correlations among variables. Another method for infusing established physical properties into an ML architecture is to assign physical significance to some neurons in the network by incorporating physical intermediate variables into the network (see Fig. 6). Daw et al. (2020), for example, implemented these two approaches in LSTM for lake temperature modeling by introducing physics-informed connections among neurons in the network. Furthermore, they added physical significance to some of the network's neurons by calculating physical intermediate elements along the network's route hence creating a unique physics-guided LSTM structure. Another related method is to set some of the neurons' weights to physically meaningful values or parameters and make them constant throughout the ML training process.

Another point of view, however, tackles the challenge of embedding prior knowledge in an ML from a different perspective. Currently, rather than creating a specific model structure that indirectly imposes prior knowledge, endeavors are being made to enforce physics-constrained learning by encoding appropriate partial differential equations (PDE) into the loss function of ML approximations utilizing methods such as automated differentiation, (see for example Sirignano and Spiliopoulos, 2018; Raissi et al., 2019; Y. Zhu et al., 2019; Geneva and Zabaras, 2020). This approach comprises a double learning process, where an algorithm learns to concurrently match the observed data and to produce outputs that approximately meet a specific set of physical laws such as energy and mass conservation.

Constraining the optimization of data-driven models is a natural way to include domain knowledge in the process of learning, and by doing so physically implausible predictions can be avoided. For problems where linear equations can be used to describe physics-based constraints, integrating the latter in existing constrained optimization formulations is straightforward. However, many agricultural problems involve complex non-linear, and dynamic processes. For instance, simulating crop yield as a function of numerous inputs such as irrigation timing and amount, plant cultivar, and soil hydraulic properties have constraints

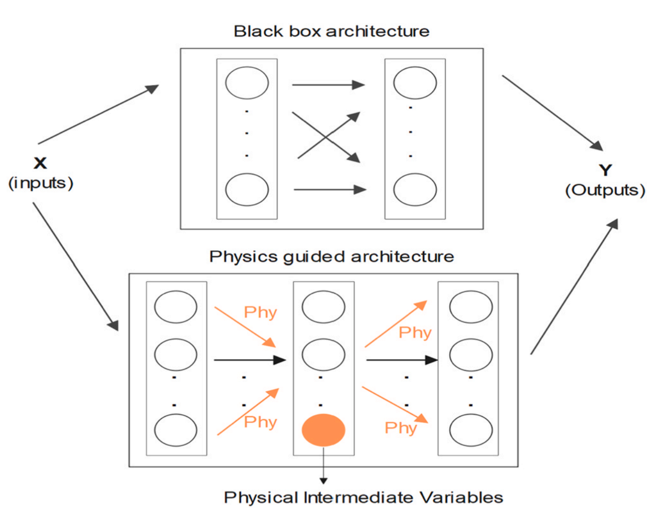

Fig. 6. Physics guided architecture of ML (adapted from Daw et al., 2020).

that are expressed in complex formulations such as non-linear relationships between variables or PDE which are not effectively treated by conventional constrained optimization approaches. In this case, it is crucial to form constrained optimization methods that can use common PDE forms and can handle the non-linearity of physical variables.

6. Discussion

This paper reviewed 16 studies that were relevant to the application of ML in supporting irrigation decision-making. The goal was to identify ML algorithms commonly used to build systems supporting irrigation decision making, analyze developed systems based on these algorithms and discuss their performance and limitations as ways to improve irrigation water use management. We observed that the input data for soil and weather-related parameters used to train the developed models were obtained mainly from sensors distributed across the study areas. Therefore, wireless sensor networks and IoT technologies have played an important role in enabling the application of ML models in irrigation water use modeling.

However, careful selection of input features is crucial to obtain a good model performance. While it is important to critically select the input data, it was demonstrated by Goldstein et al. (2018) and Jimenez et al. (2021), that having many input variables to predict irrigation demand does not imply having a better model. More than half of the reviewed papers used simple ML algorithms such as LR, RF, SVM, DT, etc. which sometimes do not accurately simulate the relationship between the plant and its environment, especially for high-dimensional datasets. As stated by Goldstein et al. (2018), Torres-Sanchez et al. (2020), and Sidhu et al. (2020), due to the non-linear relationships among soil, weather variables, and crop water requirement, linear regression models do not perform well because they cannot accurately learn the underlying functions and data patterns that relate the input to output process. Therefore, in crop water requirement prediction, the algorithms that can simulate the non-linearity between the inputs and outputs are recommended. Ensemble learning methods such as Adaboost and GBRT provide a way to improve the predictability of ML models and have demonstrated good performance in predicting irrigation management over algorithms such as LR. According to the results of Alibabaei et al. (2022), LSTM networks perform better than other algorithms because they can make use of the retained information from previous time steps in the following ones to provide better predictions. Hence, they have proven to be robust and able to accurately simulate the relationship between the dependent and independent variables when data sets from new sites are used. We postulate that PGML can be a better approach to modeling the interlinkage between the soil, plant, and climate variables. Moreover, these interactions vary extensively in time, distance, and region and physical attributes can better describe these variabilities; thereby incorporating physics into the ML algorithms can justify these interlinked processes and parameter choices for which reliable values can be obtained.

On one hand, it was observed that ML approaches have shown the potential to support irrigation decision-making and reduce the time spent by agronomists on analyzing a large quantity of data to determine irrigation water requirements. On the other hand, ML models require a wide range of data in space and time which are often not readily available. To overcome this challenge, additional data sources were discussed and data augmentation methods such as GAN are proposed as approaches to generate new data. However, both data and model uncertainty should be addressed in data-driven crop water demand modeling. We observed that there is a lack of both aleatoric and epistemic uncertainties assessments in all the surveyed studies, yet these uncertainties affect considerably the simulation results and, subsequently. the reliability of the developed systems. Additionally, there is still a lack of incorporating agronomists' knowledge in the irrigation decision-making process. To bridge this gap, human-in-the-loop approaches to enable interactions between farmers and ML models should be explored in the future. Lastly, to widen the adoption of the ML-based irrigation decision-making systems, we recommend the implementation of FAIR principles in irrigation modeling as has been the case for other disciplines such as ecological and environmental modeling (Petzold et al., 2019; Quay et al., 2021) and water science simulations (Crystal-Ornelas et al., 2022; Cannon et al., 2022). For ML-based irrigation decision-making models to have an impact on water sustainability and management, data and the resulting tools must be openly accessible, easily findable, and reusable.

7. Conclusions

This review highlights the outcomes of ML based irrigation scheduling models applied in different geographical, farm and crop settings. We discussed the current state-of-the-art models and proposed several approaches that could complement data-driven irrigation modelling. To summarize, the reviewed studies provided promising outcomes for irrigation scheduling. As we make progress in data-driven simulation, we expect to gain a new understanding of the relative importance of the choice of different neural network architecture and their workflow and structure that drive decisions and actions. The review analyses presented herein are intended to provide a basis for data-driven irrigation scheduling applications across different environmental and geographical settings. Further research is, however, needed to ensure the effectiveness of developed ML models and their applicability in farms with different soil, weather, and crop characteristics. Recognizing a growing enthusiasm for data-driven modeling applications in irrigation scheduling, we anticipate growth on several facades: (i) a better ML model capable of forecasting irrigation water use in advance, (ii) a better benchmarking irrigation model to compare with the data-driven approaches, and (iii) a better pre-processing approach for data and error estimation methods and the ways to leverage human-in-the-loop in the ML modeling system. As always, we welcome discussions with irrigation research communities interested in data-driven and other related ML modeling development to address irrigation hydrology challenges.

CRediT authorship contribution statement

Umutoni Lisa: Writing – original draft, Conceptualization. **Samadi Vidya:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data availability

this is a review paper so no research data is generated or used.

Acknowledgements

The authors appreciate the funding support from the Southern Sustainable Agriculture Research and Education (SARE; grant # 2022001722) graduate fellowship program as well as USDA-NIFA (garnt # 2023000603). Any opinions, findings, and discussions expressed in this study are those of the authors and do not necessarily reflect the views of the SARE and USDA-NIFA.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: techniques, applications and challenges. Inf. Fusion 76, 243–297. https://doi.org/10.1016/J.INFFUS.2021.05.008.

Adeyemi, O., Grove, I., Peets, S., Domun, Y., Norton, T., 2018. Dynamic neural network modelling of soil moisture content for predictive irrigation scheduling, 2018, Vol. 18, 3408 Sensors 18 (10), 3408. https://doi.org/10.3390/S18103408.

Alibabaei, K., Gaspar, P.D., Assunção, E., Alirezazadeh, S., Lima, T.M., 2022. Irrigation optimization with a deep reinforcement learning model: Case study on a site in Portugal. Agric. Water Manag. 263, 107480 https://doi.org/10.1016/J.AGWAT.2022.107480.

Allen, R.G., Pereira, L.S., Smith, M., Raes, D., Wright, J.L., 2005. FAO-56 dual crop coefficient method for estimating evaporation from soil and application extensions. J. Irrig. Drain. Eng. 131 (1), 2–13.

Amit, Y., Geman, D., 1997. Shape quantization and recognition with randomized trees. Neural Comput. 9 (7), 1545–1588. https://doi.org/10.1162/NECO.1997.9.7.1545.

Boardman, M., & Butcher, F. (2019). An exploration of maintaining human control in AI enabled systems and the challenges of achieving it. *In Workshop on Big Data Challenge-Situation Awareness and Decision Support. Brussels: North Atlantic Treaty Organization Science and Technology Organization. Porton Down: Dstl Porton Down.* https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-178/MP-IST-178–07.pdf.

Boeckhout, M., Zielhuis, G., Bredenoord, A.L., 2018. The FAIR guiding principles for data stewardship: fair enough? Eur. J. Hum. Genet. 26 (7), 931–936 https://www.nature.com/articles/s41431-018-0160-0.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Cannon, M., Kelly, A., Freeman, C., 2022. Implementing an Open & FAIR data sharing policy—A case study in the earth and environmental sciences. Learn. Publ. 35 (1), 56–66.

Cap, Q.H., Uga, H., Kagiwada, S., Iyatomi, H., 2022. LeafGAN: an effective data augmentation method for practical plant disease diagnosis. IEEE Trans. Autom. Sci. Eng. 19 (2) https://doi.org/10.1109/TASE.2020.3041499.

Car, N.J., 2018. USING decision models to enable better irrigation Decision Support Systems. Comput. Electron. Agric. 152, 290–301. https://doi.org/10.1016/J.COMPAG.2018.07.024.

Carvalho, D.V., Pereira, E.M., Cardoso, J.S., 2019. Machine learning interpretability: a survey on methods and metrics. *Electron. (Switz.)* Vol. 8 (Issue 8). https://doi.org/10.3390/electronics8080832.

Chen, M., Cui, Y., Wang, X., Xie, H., Liu, F., Luo, T., Zheng, S., Luo, Y., 2021. A reinforcement learning approach to irrigation decision-making for rice using weather forecasts. Agric. Water Manag. 250, 106838 https://doi.org/10.1016/J.AGWAT.2021.106838.

Chen, Y., Hsieh, W., & Y. K.–. (2021). An ensemble learning model for agricultural irrigation prediction. *Ieeexplore.Ieee.Org.* https://ieeexplore.ieee.org/abstract/document/9333852/.

Committee on Best Practices for a Future Open Code Policy for NASA Space Science, S. S. B. & D. on E. and P. S. (2018). *Open Source Software Policy Options for NASA Earth and Space Sciences.* https://books.google.com/books?hl=en&lr=&id=BP6EDwAAQBAJ&oi=fnd&pg=PR1&dq=Open+Source+Software+Policy+Options+for+NASA+Earth+and+Space+Sciences+(2018)&ots=syMPdkptRX&sig=-XTycsx-OXtH_-3lNVWJ6zBhI7c.

Crystal-Ornelas, R., Varadharajan, C., O'Ryan, D., Beilsmith, K., Bond-Lamberty, B., Boye, K., Burrus, M., Cholia, S., Christianson, D.S., Crow, M., Damerow, J., 2022. Enabling FAIR data in Earth and environmental science with community-centric (meta) data reporting formats. Sci. data 9 (1), 700.

Dahane, A., Benameur, R., Kechar, B., & Benyamina, A. (2020). *An IoT Based Smart Farming System Using Machine Learning; An IoT Based Smart Farming System Using Machine Learning.* https://doi.org/10.1109/ISNCC49221.2020.9297341.

Daw, A., Thomas, R.Q., Carey, C.C., Read, J.S., Appling, A.P., & Karpatne, A. (2020). Physics-guided architecture (pga) of neural networks for quantifying uncertainty in lake temperature modeling. *Proceedings of the 2020 SIAM International Conference on Data Mining, SDM 2020*, 532–540. https://doi.org/10.1137/1.9781611976236.60.

Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2). https://www.sciencedirect.com/science/article/pii/S0167473008000556.

Dibike, Y.B., Velickov, S., Solomatine, D., Abbott, M.B., 2001. Model induction with support vector machines: introduction and applications. J. Comput. Civ. Eng. 15 (3) https://doi.org/10.1061/(asce)0887-3801(2001)15:3(208).

Divyanth, L.G., Guru, D.S., Soni, P., Machavaram, R., Nadimi, M., Paliwal, J., 2022. Image-to-image translation-based data augmentation for improving crop/weed classification models for precision agriculture applications. Algorithms 15 (11). https://doi.org/10.3390/a15110401.

Donratanapat, N., Samadi, S., Vidal, J., 2020. A national scale big data analytics pipeline to assess the potential impacts of flooding on critical infrastructures and communities. Environ. Model. Softw. 133 (104828) https://www.sciencedirect.com/science/article/pii/S1364815220308859.

Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D., 1987. Hybrid monte carlo. Phys. Lett. B 195 (2), 216–222 https://www.sciencedirect.com/science/article/pii/S1364815220308859.

El Bilali, A., Abdeslam, T., Ayoub, N., Lamane, H., Ezzaouini, M.A., Elbeltagi, A., 2023. An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation. J. Environ. Manag. 327 https://doi.org/10.1016/j.jenvman.2022.116890.

Elijah, O., Rahman, T.A., Orikumhi, I., Leow, C.Y., Hindia, M.H.D.N., 2018. An overview of internet of things (IoT) and data analytics in agriculture: benefits and challenges. IEEE Internet Things J. 5 (5), 3758–3773. https://doi.org/10.1109/JIOT.2018.2844296.

Espejo-Garcia, B., Panoutsopoulos, H., Anastasiou, E., Rodríguez-Rigueiroz, F.J., Fountas, S., 2023. Top-tuning on transformers and data augmentation transferring for boosting the performance of weed identification. Comput. Electron. Agric. 211 https://doi.org/10.1016/j.compag.2023.108055.

Fawakherji, M., Potena, C., Prevedello, I., Pretto, A., Bloisi, D.D., & Nardi, D. (2020). Data Augmentation Using GANs for Crop/Weed Segmentation in Precision Farming. *CCTA 2020 - 4th IEEE Conference on Control Technology and Applications*. https://doi.org/10.1109/CCTA41146.2020.9206297.

García, L., Parra, L., Jimenez, J., Lloret, J., Lorenz, P., 2020. IoT-based smart irrigation systems: an overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture. Sensors 20 (4). https://doi.org/10.3390/s20041042.

Geneva, N., Zabaras, N., 2020. Modeling the dynamics of PDE systems with physics-constrained deep auto-regressive networks. J. Comput. Phys. 403 https://doi.org/10.1016/J.JCP.2019.109056.

Gil, Y., David, C.H., Demir, I., Essawy, B.T., Fulweiler, R.W., Goodall, J.L., Karlstrom, L., Lee, H., Mills, H.J., Oh, J.H., Pierce, S.A., Pope, A., Tzeng, M.W., Villamizar, S.R., Yu, X., 2016. Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. Earth Space Sci. 3 (10), 388–415. https://doi.org/10.1002/2015EA000136.

Gill, M.K., Asefa, T., Kemblowski, M.W., McKee, M., 2006. Soil moisture prediction using support vector machines. J. Am. Water Resour. Assoc. 42 (4) https://doi.org/10.1111/j.1752-1688.2006.tb04512.x.

Glória, A., Cardoso, J., Sebastião, P., 2021. Sustainable irrigation system for farming supported by machine learning and real-time sensor data, 2021, Vol. 21, 3079 Sensors 21 (9), 3079. https://doi.org/10.3390/S21093079.

Goap, A., Sharma, D., Shukla, A.K., Rama Krishna, C., 2018. An IoT based smart irrigation management system using Machine learning and open source technologies. Comput. Electron. Agric. 155, 41–49. https://doi.org/10.1016/J.COMPAG.2018.09.040.

Goldstein, A., Fink, L., Meitin, A., Bohadana, S., Lutenberg, O., Ravid, G., 2018. Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge. Precis. Agric. 19 (3), 421–444. https://doi.org/10.1007/S11119-017-9527-4.

Gu, Z., Qi, Z., Burghate, R., Yuan, S., Jiao, X., Xu, J., 2020. Irrigation scheduling approaches and applications: a review. J. Irrig. Drain. Eng. 146 (6) https://doi.org/10.1061/(asce)ir.1943-4774.0001464.

Gumiere, S.J., Camporese, M., Botto, A., Lafond, J.A., Paniconi, C., Gallichand, J., Rousseau, A.N., 2020. Machine Learning vs. Physics-Based Modeling for Real-Time Irrigation Management. Front. Water 2. https://doi.org/10.3389/frwa.2020.00008.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/NECO.1997.9.8.1735.

Huang, X., Ouyang, S., Ma -, S., Sharma, P., Jain, J.K., Kallaa, P., Sun, A.Y., Scanlon, B.R., 2019. How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. Environ. Res. Lett. 14 (7), 073001 https://doi.org/10.1088/1748-9326/AB1B7D.

Huang, Y., Li, R., Wei, X., Wang, Z., Ge, T., Qiao, X., 2022. Evaluating data augmentation effects on the recognition of sugarcane leaf spot. Agric. (Switz.) 12 (12). https://doi.org/10.3390/agriculture12121997.

Hunsaker, D.J., Barnes, E.M., Clarke, T.R., Fitzgerald, G.J., Pinter, P.J., 2005. Cotton irrigation scheduling using remotely sensed and FAO-56 basal crop coefficients. Trans. ASAE 48 (4), 1395–1407. https://doi.org/10.13031/2013.19197.

Jamroen, C., Komkum, P., Fongkerd, C., Krongpha, W., 2020. An intelligent irrigation scheduling system using low-cost wireless sensor network toward sustainable and precision agriculture. IEEE Access 8, 172756–172769. https://doi.org/10.1109/ACCESS.2020.3025590.

Jimenez, A.F., Cardenas, P.F., Jimenez, F., Canales, A., López, A., 2020. A cyber-physical intelligent agent for irrigation scheduling in horticultural crops. Comput. Electron. Agric. 178, 105777 https://doi.org/10.1016/J.COMPAG.2020.105777.

Jimenez, A.-F., Ortiz, B.V., Bondesan, L., Morata, G., Damianidis, D., 2021. Long short-term memory neural network for irrigation management: a case study from Southern Alabama, USA. Precis. Agric. 22 (2), 475–492. https://doi.org/10.1007/s11119-020-09753-z.

Jones, H.G., 2004. Irrigation scheduling: advantages and pitfalls of plant-based methods. J. Exp. Bot. 55 (407) https://doi.org/10.1093/jxb/erh213.

Kamienski, C., Soininen, J.-P., Taumberger, M., Dantas, R., Toscano, A., Salmon Cinotti, T., Filev Maia, R., Torre Neto, A., 2018. Smart water management platform: IoT-based precision irrigation for agriculture. Mdpi. Com. https://doi.org/10.3390/s19020276.

Karandish, F., Šimůnek, J., 2016. A comparison of numerical and machine-learning modeling of soil water content with limited input data. J. Hydrol. 543 https://doi.org/10.1016/j.jhydrol.2016.11.007.

Karasekreter, N., Başçiftçi, F., Fidan, U., 2013. A new suggestion for an irrigation schedule with an artificial neural network. J. Exp. Theor. Artif. Intell. 25 (1), 93–104. https://doi.org/10.1080/0952813X.2012.680071.

Kashyap, P.K., Kumar, S., Jaiswal, A., Prasad, M., Gandomi, A.H., 2021. Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network. IEEE Sens. J. 21 (16), 17479–17491. https://doi.org/10.1109/JSEN.2021.3069266.

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., Wortman Vaughan, J., 2020. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. Conf. Hum. Factors Comput. Syst. - Proc. https://doi.org/10.1145/3313831.3376219.

Kavitha, S., Varuna, S., & R. Ramya. (2016). A comparative analysis on linear regression and support vector regression. *Online International Conference on Green Engineering and Technologies (IC-GET)*. https://ieeexplore.ieee.org/abstract/document/7916627/.

Kecman, V., 2005. Support Vector Machines – An Introduction. Support Vector Mach.: Theory Appl. 1–47. https://doi.org/10.1007/10984697_1.

Kendall, A., Gal, Y., 2017. What uncertainties do we need in Bayesian deep learning for computer vision? Adv. Neural Inf. Process. Syst. 5575–5585. *2017-December*.

Koymans, M.R., van Hinsbergen, D.J.J., Pastor-Galán, D., Vaes, B., Langereis, C.G., 2020. Towards FAIR paleomagnetic data management through paleomagnetism.org 2.0. Geochem., Geophys., Geosystems 21 (2), e2019GC008838. https://doi.org/10.1029/2019GC008838.

Kumar, A., Surendra, A., & Mohan, H. (2017). Internet of things based smart irrigation using regression algorithm. *017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 1652–1657. https://ieeexplore.ieee.org/abstract/document/8342819/.

Lakshminarayanan, B., Pritzel, A., & Deepmind, C.B. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Proceedings.Neurips.Cc*. https://proceedings.neurips.cc/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html.

Li, P., Zha, Y., Shi, L., Tso, C.H.M., Zhang, Y., Zeng, W., 2020. Comparison of the use of a physical-based model with data assimilation and machine learning methods for simulating soil water dynamics. J. Hydrol. 584 https://doi.org/10.1016/j.jhydrol.2020.124692.

Liou, Y.A., Liu, S.F., Wang, W.J., 2001. Retrieving soil moisture from simulated brightness temperatures by a neural network. IEEE Trans. Geosci. Remote Sens. 39 (8) https://doi.org/10.1109/36.942544.

Liu, L., Ismail, M., Wang, Y., Lin, W.S., 2021. Internet of things based smart irrigation control system for paddy field. AGRIVITA, J. Agric. Sci. 43 (2), 378–389. https://doi.org/10.17503/agrivita.

Mahesh, B., 2018. Machine learning algorithms-a review. Int. J. Sci. Res. https://doi.org/10.21275/ART20203995.

Maier, N., Dietrich, J., 2016. Using SWAT for strategic planning of basin scale irrigation control policies: a case study from a humid region in Northern Germany. Water Resour. Manag. 30 (9) https://doi.org/10.1007/s11269-016-1348-0.

Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. Artif. Intell. Vol. 267 https://doi.org/10.1016/j.artint.2018.07.007.

Mitchell, T. (2007). *Machine learning* (Vol. 1). https://profs.info.uaic.ro/~ciortuz/SLIDES/2017s/ml0.pdf.

Moraffah, R., Karami, M., Guo, R., Raglin, A., Liu, H., 2020. Causal interpretability for machine learning - problems, methods and evaluation. ACM SIGKDD Explor. Newsl. 22 (1) https://doi.org/10.1145/3400051.3400058.

Nachankar, P.J., Somani, M.G., Singh, D.M., Katkar, S.N., 2018. IOT in agriculture. Int. Res. J. Eng. Technol. (www.irjet.net).

National Academies of Sciences, E. and M. (2021). Human-AI Teaming: State-of-the-Art and Research Needs. *Human-AI Teaming: State-of-the-Art and Research Needs* (2022), 1–126. https://doi.org/10.17226/26355.

Navarro-Hellín, H., Martinez-del-Rincon, J., Domingo-Miguel, R., Soto-Valles, F., Torres-Sánchez, R., 2016. A decision support system for managing irrigation in agriculture. Comput. Electron. Agric. 124, 121–131.

Nawandar, N.K., Satpute, V.R., 2019. IoT based low cost and intelligent module for smart irrigation system. Comput. Electron. Agric. 162, 979–990. https://doi.org/10.1016/J.COMPAG.2019.05.027.

Nemali, K.S., van Iersel, M.W., 2006. An automated system for controlling drought stress and irrigation in potted plants. Sci. Hortic. 110 (3), 292–297. https://doi.org/10.1016/J.SCIENTA.2006.07.009.

Oden Technologies. (2023). What is Model Training | Oden Technologies. https://oden.io/glossary/model-training/.

Petzold, A., Asmi, A., Vermeulen, A., Pappalardo, G., Bailo, D., Schaap, D., Glaves, H.M., Bundke, U. and Zhao, Z., 2019, September. ENVRI-FAIR-interoperable environmental FAIR data and services for society, innovation and research. In 2019 15th International Conference on eScience (eScience) (pp. 277–280). IEEE.

Pfeiffer III, J., Neville, J., & Bennett, P. (2015). Overcoming relational learning biases to accurately predict preferences in large scale networks. *In Proceedings of the 24th International Conference on World Wide Web*, 853–863. https://doi.org/10.1145/2736277.2741668.

Pierce, F., & Nowak, P. (1999). Aspects of precision agriculture. *Advances in Agronomy*, 1–85. https://www.sciencedirect.com/science/article/pii/S0065211308605131.

Perea, R.G., Poyato, E.C., Montesinos, P., Díaz, J.R., 2019. Prediction of irrigation event occurrence at farm level using optimal decision trees. Comput. Electron. Agric. 157 https://doi.org/10.1016/j.compag.2018.12.043.

Quay, A.N., Fiske, P.S., Mauter, M.S., 2021. Recommendations for advancing FAIR and open data standards in the water treatment community. ACS EST Eng. 2 (3), 337–346.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. https://doi.org/10.1016/J.JCP.2018.10.045.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Model-agnostic interpretability of machine learning. *ArXiv Preprint*. ArXiv 1606, *05386*.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* Vol. 1 (Issue 5) https://doi.org/10.1038/s42256-019-0048-x.

Sabrina, F., Sohail, S., Farid, F., Jahan, S., Ahamed, F., Gordon, S., 2022. An interpretable artificial intelligence based smart agriculture system. Comput., Mater. Contin. 72 (2) https://doi.org/10.32604/cmc.2022.026363.

Saggi, M.K., Jain, S., 2022. A survey towards decision support system on smart irrigation scheduling using machine learning approaches. Arch. Comput. Methods Eng. https://doi.org/10.1007/s11831-022-09746-3.

Samadi, S., Pourreza-Bilondi, M., Wilson, C.A.M.E., Hitchcock, D.B., 2020. Bayesian model averaging with fixed and flexible priors: theory, concepts, and calibration experiments for rainfall-runoff modeling. J. Adv. Model. Earth Syst. 12 (7) https://doi.org/10.1029/2019MS001924.

Sharma, A., Jain, A., Gupta, P., Chowdary, V., 2021. Machine learning applications for precision agriculture: a comprehensive review. IEEE Access 9, 4843–4873. https://doi.org/10.1109/ACCESS.2020.3048415.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.J., Ganguly, S., Hsu, K. L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., Tsai, W.P., 2018. HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. Hydrol. Earth Syst. Sci. 22 (11), 5639–5656. https://doi.org/10.5194/HESS-22-5639-2018.

Shrestha, N.K., Shukla, S., 2015. Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. Agric. For. Meteorol. 200 https://doi.org/10.1016/j.agrformet.2014.09.025.

Sidhu, R.K., Kumar, R., Rana, P.S., 2020. Machine learning based crop water demand forecasting using minimum climatological data. Multimed. Tools Appl. 79 (19–20), 13109–13124. https://doi.org/10.1007/S11042-019-08533-W.

Sihi, D., Dari, B., Kuruvila, A.P., Jha, G., Basu, K., 2022. Explainable machine learning approach quantified the long-term (1981–2015) impact of climate and soil properties on yields of major agricultural crops across CONUS. Front. Sustain. Food Syst. 6 https://doi.org/10.3389/fsufs.2022.847892.

Sirignano, J., Spiliopoulos, K., 2018. DGM: a deep learning algorithm for solving partial differential equations. J. Comput. Phys. 375, 1339–1364. https://doi.org/10.1016/J.JCP.2018.08.029.

Su, D., Kong, H., Qiao, Y., Sukkarieh, S., 2021. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. Comput. Electron. Agric. 190 https://doi.org/10.1016/j.compag.2021.106418.

Sun, C., Ren, L., 2014. Assessing crop yield and crop water productivity and optimizing irrigation scheduling of winter wheat and summer maize in the Haihe plain using SWAT model. Hydrol. Process. 28 (4) https://doi.org/10.1002/hyp.9759.

Sun, L., Yang, Y., Hu, J., Porter, D., & Marek, T. (2017). Reinforcement learning control for water-efficient agricultural irrigation. In *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, 1334–1341. https://ieeexplore.ieee.org/abstract/document/8367433/.

Tabas, S.S., Samadi, S., 2022. Variational Bayesian dropout with a Gaussian prior for recurrent neural networks application in rainfall-runoff modeling. Environ. Res. Lett. 17 (6).

Tace, Y., Tabaa, M., Elfilali, S., Leghris, C., Bensag, H., Renault, E., 2022. Smart irrigation system based on IoT and machine learning. Energy Rep. 8 https://doi.org/10.1016/j.egyr.2022.07.088.

Torres-Sanchez, R., Navarro-Hellin, H., Guillamon-Frutos, A., San-Segundo, R., Ruiz-Abellón, M.C., Domingo-Miguel, R., 2020. A Decision Support System for Irrigation Management: Analysis and Implementation of Different Learning Techniques, 2020, Vol. 12, 548 Water 12 (2), 548. https://doi.org/10.3390/W12020548.

Tseng, D., Wang, D., Chen, C., & Miller, L. (2018). Towards automating precision irrigation: Deep learning to infer local soil moisture conditions from synthetic aerial agricultural images. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, 284–291. https://ieeexplore.ieee.org/abstract/document/8560431/.

2017 USDA. (2017). *2017 Census of Agriculture*. (United States Department of Agriculture - National Agricultural Statistics Service). https://www.nass.usda.gov/Publications/AgCensus/2017/index.php.

Vellidis, G., V. Liakos, W. Porter, M. Tucker, X. Liang. 2016. A Dynamic Variable Rate Irrigation System. In Proceedings of the 13th International Conference on Precision Agriculture July 31 – August 3, 2016, St. Louis, Missouri, USA. International Society of Precision Agriculture, Monticello, IL.

Wagner, A., Waite, L., Wierzba, M., Hoffstaedter, F., 2022. FAIRly big: a framework for computationally reproducible processing of large-scale data. Sci. Data 9 (1), 80 https://www.nature.com/articles/s41597-022-01163-2.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *2016 3:1* Sci. Data 3 (1), 1–9. https://doi.org/10.1038/sdata.2016.18.

Windheuser, L., Karanjit, R., Pally, R., Samadi, S., Hubig, N.C., 2023. An end-to-end flood stage prediction system using deep neural networks. Earth Space Sci. 10 (1) https://doi.org/10.1029/2022EA002385.

Wolanin, A., Mateo-Garciá, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., Guanter, L., 2020. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. Environ. Res. Lett. 15 (2) https://doi.org/10.1088/1748-9326/ab68ac.

Wu, Q., Chen, Y., Meng, J., 2020. Dcgan-based data augmentation for tomato leaf disease identification. IEEE Access 8. https://doi.org/10.1109/ACCESS.2020.2997001.

Yang, Y., Hu, J., Porter, D., Marek, T., Heflin, K., Kong, H., Sun, L., 2020. Deep reinforcement learning-based irrigation scheduling. Trans. ASABE 63 (3), 549–556. https://doi.org/10.13031/TRANS.13633.

Zhang, J., Guan, K., Peng, B., Jiang, C., Zhou, W., Yang, Y., Pan, M., Franz, T., Heeren, D. M., Rudnick, D.R., 2021. Challenges and opportunities in precision irrigation decision-support systems for center pivots. Environ. Res. Lett. 16 (5), 053003 https://iopscience.iop.org/article/10.1088/1748-9326/abe436/meta.

Zhao, X., Zhang, L., Zhu, G., Cheng, C., He, J., Traore, S., Singh, V.P., 2023. Exploring interpretable and non-interpretable machine learning models for estimating winter wheat evapotranspiration using particle swarm optimization with limited climatic data. Comput. Electron. Agric. 212 https://doi.org/10.1016/j.compag.2023.108140.

Zhou, L., Pan, S., Wang, J., Vasilakos, A.V., 2017. Machine learning on big data: Opportunities and challenges. Neurocomputing 237, 350–361. https://doi.org/10.1016/J.NEUCOM.2017.01.026.

Zhou, N., 2020. Intelligent control of agricultural irrigation based on reinforcement learning. J. Phys.: Conf. Ser. 1601, 52031. https://doi.org/10.1088/1742-6596/1601/5/052031.

Zhu, Y., Zabaras, N., Koutsourelakis, P.S., Perdikaris, P., 2019. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. J. Comput. Phys. 394, 56–81. https://doi.org/10.1016/J.JCP.2019.05.024.

Zhu, Z., Wulder, M., Roy, D., Woodcock, C., 2019. Benefits of the free and open Landsat data policy. Remote Sens. Environ. 224, 382–385 https://www.sciencedirect.com/science/article/pii/S0034425719300719.