DOI: 10.5958/0974-4444.2024.00010.7

Vol. 53 (Special Issue), pp. 106–113 Indian Journal of Nematology March, 2024

Advancements in Nematode Management: Exploring Machine Learning in Precision Agriculture

ABHISHEK MUKHERJEE[1*], SRISTI DAS[1], SUVASRI DUTTA[1] and SANDIP MONDAL[2]

[1]*Agricultural and Ecological Research Unit, Indian Statistical Institute, Giridih, Jharkhand - 815 301, India*;
[2]*Department of Plant Pathology, The Ohio State University, Columbus, Ohio, USA, 43210*

[*]Principal Author E-mail: abhi.mukh@yahoo.com; ORCID ID: 0000-0002-5277-1363

ABSTRACT: Agriculture serves as the backbone of India's economy. The ongoing rise in global population, coupled with climate change and rapid urbanization, has significantly impacted agriculture. Maximizing crop yields is imperative to ensure global food security. Precision agriculture, or the integration of computation into agricultural practices is not a recent concept. Machine learning (ML) algorithms have long been employed in tasks such as vegetation analysis, crop modelling, and yield management. However, the application of ML in agricultural pest management is currently gaining attention, as effective pest control necessitates accurate pest identification - a challenging task. Automating this process holds promise for more efficient disease management. Various ML algorithms have been utilized in precision agriculture, and now, some are being explored for the identification and management of plant parasitic nematodes, which pose a significant threat to agricultural crops worldwide. This article provides an overview of the algorithms utilized and the databases developed to facilitate the efficient identification of plant parasitic nematodes.

Keywords: Artificial intelligence, machine learning, nematode identification

INTRODUCTION

Climate change over the last few years has been of huge concern for scientists worldwide. Plant growth and productivity largely depend upon the climatic conditions of a region. Temperature, water availability, and soil conditions immensely influence agricultural output (Anwar *et al*., 2013). On the other hand, the human population is ever-growing and is estimated to reach 9.1 billion by 2050. To feed this population, the demand for food supply will increase by about 70 per cent (Sharma *et al*., 2021).

In the present scenario where the availability of arable land is on the decline due to fast urbanization, increasing crop productivity is a real challenge. Unpredictable weather, climatic fluctuations, unplanned agricultural systems, along with the lack of proper pest management, contribute to the reduction in domestic food production (Jansson and Hofmockel, 2020; Sharma *et al*., 2021). Agriculture contributes largely to the economic development of any country. Such hindrances in food production adversely affect economic growth as well. These drawbacks need to be addressed urgently. There is an immediate need for optimization of agricultural practices which will ensure sustainability, yield maximization, as well as environmental safety. The concept of precision agriculture aims to address these challenges. In the past few years, initiatives have been taken to improve crop production by the use of artificial intelligence (AI). AI is a new and emerging tool with multi-dimensional applications. The use of machine learning (ML) technologies allows more precision in agricultural practices, thus reducing or bypassing the impact of both biotic and abiotic factors (Eli-Chukwu, 2019; Benos *et al*., 2021).

Machine learning comes under the broad segment of AI and involves analyzing raw data, extracting information from them, and generating predictions and hypotheses based on them. In the present-day situation, precision agriculture might be a potent solution for the existing conundrums. Experts in current agro-industries are trying to explore their theories largely, thus helping to achieve more accurate predictions. New-age agriculture has the scope of employing ML algorithms for disease forecasting, climatic prediction, optimization of water usage, nutrient availability, and so on (Xing and Wang, 2017; Elavarasan *et al*., 2018). The general method of pest management is by the use of chemical pesticides in fields, which are expensive and have consequences on the environment. The use of ML in agriculture allows the precise application of pesticides in accordance with time, dosage, and the nature of pathogens, thus paving the way for a more cost-effective pest management strategy (Elavarasan *et al*., 2018). The use of computation in agriculture dates back to the 1980s and has been used for the management of pests, diseases, prediction of soil, rainfall, vegetation, weed management, and control of crop products and storage (Banerjee *et al*., 2018).

Among agricultural pests, plant parasitic nematodes (PPNs) are a major threat to crop production. Nematodes are one of the most diverse groups of metazoans; still, less than 0.01 per cent of the species have been described to date. Of these, about 4100 are plant parasitic. While free-living nematodes are significant in nutrient cycling and as indicators of soil health, PPNs cause a huge loss of crops including rice, tomato, brinjal, chilli, potato, soybean, and others (Jones *et al*., 2013). PPNs are soil-borne pathogens and cause an annual yield loss of about USD 157 billion, of which, India alone incurs a loss of \$1.58 billion (Kumar *et al*., 2020). Nematodes are diverse and are found in all types of soil systems. PPNs can be terrestrial or aquatic, ectoparasites, endoparasites, or semi-endoparasites in nature (Mandal *et al*., 2021). Proper identification of nematodes is necessary to: (i) evaluate the importance of free-living nematodes as biotic indicators of soil health, (ii) assess the impact of PPNs in crop loss and disease forecasting. Automating the identification of nematodes is of utmost importance, considering the enormity of the task in the absence of an adequate number of skilled taxonomists (Bhat *et al*., 2022; Shabrina *et al*., 2023). Proper identification and enumeration in samples are key strategies in the control of PPNs (Ropelewska *et al*., 2023). This review summarizes the different machine learning algorithms used in agriculture and the ones already in use for nematode management.

CLASSIFICATION OF MACHINE LEARNING ALGORITHMS USED IN AGRICULTURE

Machine Learning (ML) involves the development of statistical models and algorithms that enable computers to perform tasks more efficiently without explicit programming. ML algorithms are typically categorized as follows:

Supervised learning

Supervised learning utilizes a labelled training dataset, consisting of input and output data, to train models for forecasting events. This involves generating a mapping system between input variables (X) and output variables (Y) (Cunningham *et al*., 2020). Key supervised ML algorithms include:

Decision Tree: A decision tree is utilized for both regression and classification problems. It constructs a tree-like structure based on input attributes, where each node represents an attribute and branches represent outcomes (Song and Lu, 2015; Nasteski, 2017).

Artificial Neural Networks (ANN): ANNs simulate neurons in the human brain and consist of layers of artificial neurons connected by weights. They are classified into feedback and feedforward networks,

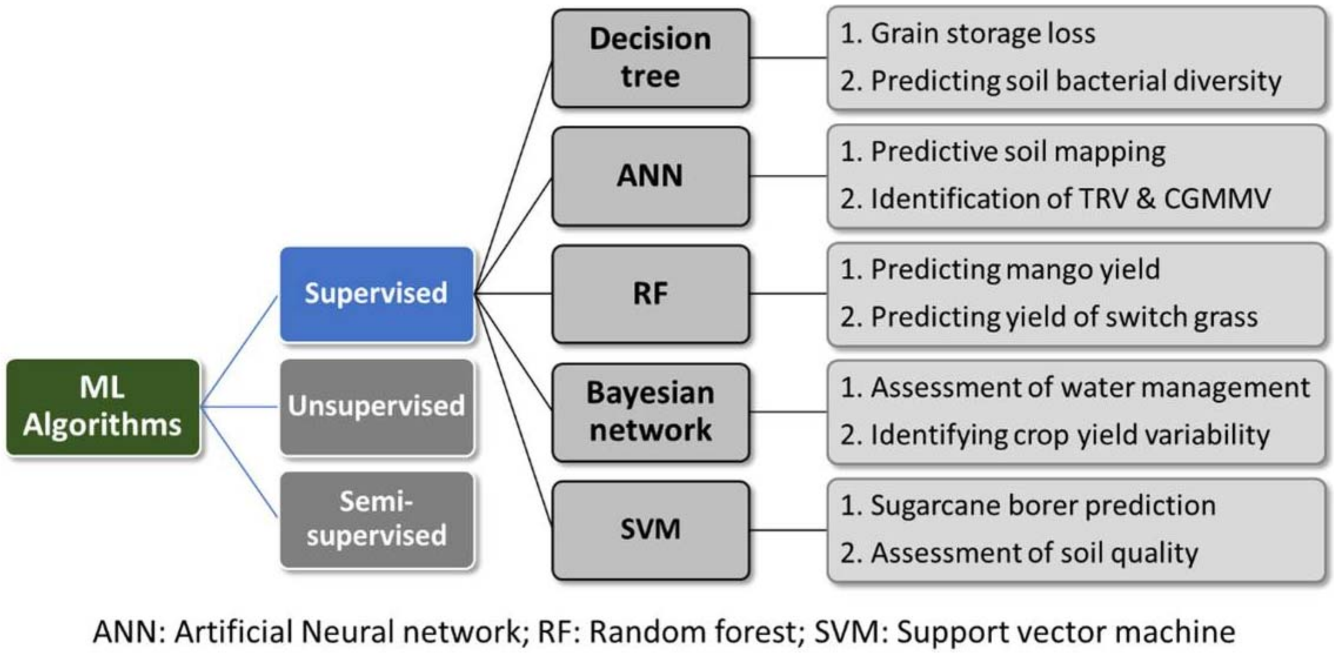


Fig. 1. An overview of supervised machine learning classification and algorithms

each with distinct characteristics (Agatonovic-Kustrin and Beresford, 2000; Hoang *et al*., 2021).

Support Vector Machine (SVM): SVM is effective for classification and regression tasks, aiming to find a hyperplane that maximizes the gap between data points for accurate classification (Zoppis *et al*., 2018; Pisner and Schnyer, 2020).

Random Forest (RF): RF utilizes an ensemble of decision trees to predict continuous variables and has applications in agriculture, particularly in crop yield forecasting (Breiman, 2001; Elavarasan *et al.*, 2018).

Bayesian Network (BN): BNs analyse interactions between variables in a dataset to generate outcomes, making them suitable for environmental and agricultural predictions (Chen and Pollino, 2012).

Unsupervised learning

Unsupervised learning algorithms generate outputs from unlabeled input data, detecting structures and patterns within datasets. Key unsupervised ML algorithms include:

Markov Chain Model (MCM): MCM records transitions in complex systems and has applications in rainfall prediction and agriculture (Sonnadara and Jayewardene, 2015).

DBSCAN: DBSCAN identifies dense regions or clusters in data, useful for crop yield prediction and plant location identification (Leroux *et al*., 2019; Shi *et al*., 2023).

K-means clustering: K-means classifies data into clusters based on similarity, with applications in various fields including agriculture (Kodinariya and Makwana, 2013).

Reinforcement Immune Algorithm (RIA): RIA mimics the human immune system and is used for resource allocation in agriculture (Elavarasan *et al.*, 2018; Jiang *et al*., 2018).

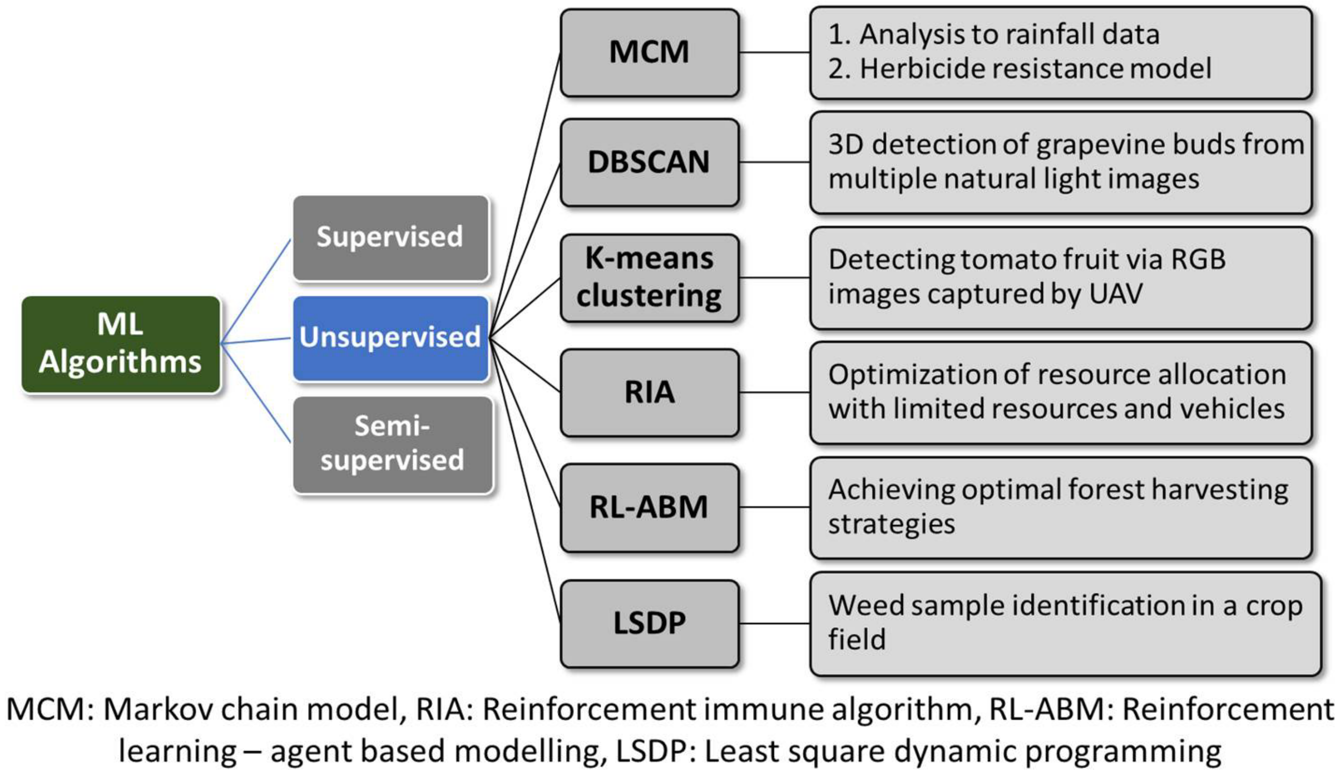


Fig. 2. An overview of unsupervised machine learning classification and algorithms

Reinforcement learning - agent-based modelling (RL-ABM): RL-ABM combines reinforcement learning with agent-based modelling for applications in disease prediction and ecosystem behaviour (Jiang *et al*., 2018; Sert *et al*., 2020).

Least square dynamic programming (LSDP): LSDP, combining statistical tools and AI, is used for weed identification in agricultural fields (Bonneau *et al*., 2014).

Semi-supervised learning

Semi-supervised learning utilizes a small amount of labelled data along with unlabelled data to train models, aiming to overcome drawbacks of both supervised and unsupervised learning. Key algorithms include:

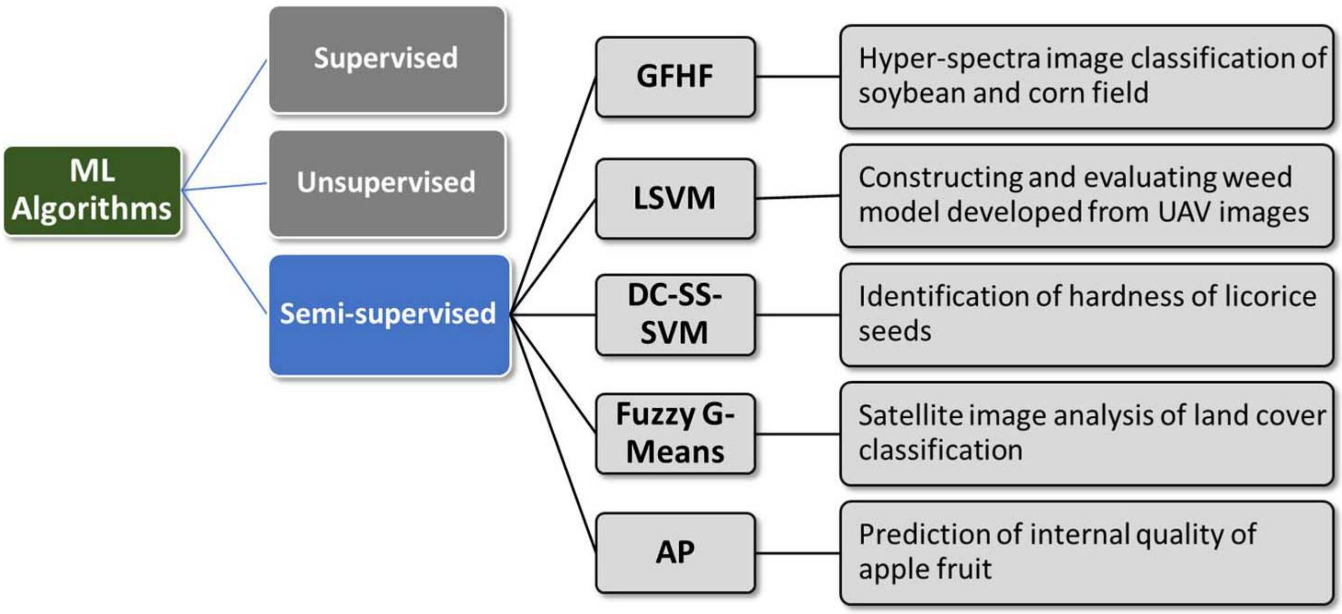
Gaussian fields and harmonic functions (GFHF): GFHF is used for classification tasks, including land cover classification in agriculture (de Sousa, 2015; Ma *et al*., 2016).

Linear Support Vector Machine (LSVM): LSVM determines vegetation indices and has applications in weed mapping using UAV imagery (Tang, 2013; Pérez-Ortiz *et al*., 2015).

Difference of convex function semi-supervised SVM (DC-SS-SVM): DC-SS-SVM is an extension of SVM for classification tasks, with applications in seed toughness determination using NIR spectroscopy (Sun and Park, 2017).

Fuzzy C-Means (FCM): FCM is used for image segmentation and land cover classification in agriculture (Nowpada *et al*., 2011; Ngo *et al*., 2021).

Affinity Propagation (AP) algorithm: AP is a clustering algorithm with applications in analysing hyperspectral data for agriculture, such as determining soluble solid content in apples (Zhu *et al*., 2013).



GFHF: Gaussian fields and harmonic functions, LSVM: Linear support vector machine, DC-SS-SVM: Difference of convex function semi supervised SVM, AP: Affinity propagation algorithm

Fig. 3. An overview of semi-supervised machine learning classification and algorithms

ML IN NEMATODE IDENTIFICATION

Plant parasitic nematodes (PPNs) pose a significant threat to agriculture, and traditional control methods are often expensive, time-consuming, and labour-intensive. Moreover, nematode infections are challenging to detect, as infected plants may exhibit general symptoms similar to those caused by other factors. Additionally, PPNs can form disease complexes with other pathogens, further complicating control efforts. Accurate identification of PPN species is crucial for effective control measures, but the declining number of skilled taxonomists presents a challenge. Automation of this task offers a promising solution in terms of accuracy, efficiency, and cost-effectiveness, with machine learning (ML) algorithms emerging as a key approach for nematode identification.

Several databases and ML-based systems have been developed for image-based species-level identification of PPNs:

NEMANet

Developed by Abade *et al*. (2022), NEMANet is a convolutional neural network (CNN) system trained on image data of nematodes damaging soybeans. CNNs, known for their ability to automatically extract features from data, are a state-of-the-art algorithm for image detection and classification. NEMANet achieved an impressive accuracy of 96.99–98.88 per cent on average, demonstrating its effectiveness in identifying PPN species.

NemaRec

NemaRec is a database designed for the identification of soil nematodes, utilizing a convolutional neural network for image processing and identification. Although user-friendly, NemaRec's identification accuracy averages only 54.7 per cent, indicating room for improvement (Qing *et al*., 2022).

NemDST

NemDST is a decision support system developed to identify and count populations of root-knot nematodes (RKNs) in soil. Utilizing a deep-learning approach based on YOLOv5 model, NemDST demonstrates high accuracy in detection, classification, and counting of nematodes (Pun *et al*., 2023).

Artificial Neural Network (ANN) system

Ropelewska *et al*. (2023) developed an ANN system for identifying cyst nematodes, achieving accuracies ranging from 83.7 to 98 per cent for different species.

Free-living marine nematodes identification

ML algorithms including Random Forest (RF), Support Vector Machine (SVM), K-nearest neighbour (KNN), and Stochastic Gradient Boosting (SGBoost) were used for identifying free-living marine nematodes. Among these, the Random Forest algorithm demonstrated the highest precision, correctly identifying 100 per cent of Sabatieria species (Brito de Jesus *et al*. 2023).

These databases and ML-based systems represent promising advancements in the field of nematode identification, offering efficient and accurate solutions for agricultural management.

CONCLUSION

Although crop losses caused by PPN attacks often go unnoticed, they are estimated to exceed 12 per cent globally across the 40 most important crops (Kumar *et al*., 2020). Effective nematode management has become a pressing concern. Presently, machine learning algorithms are being developed for precise nematode identification. Most of these tools have demonstrated high accuracy in classifying and identifying PPNs. The integration of computation into nematode identification and management holds significant potential for disease control and warrants thorough exploration. The creation of more databases is necessary to enable growers to accurately identify species and implement appropriate pesticide measures. Successful implementation of machine learning in pest management represents a significant step toward a sustainable future.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

Abade, A., Porto, L.F., Ferreira, P.A. & de Barros Vidal, F. (2022). NemaNet: A convolutional neural network model for identification of soybean nematodes. *Biosystems Engineering* **213:** 39–62.

Agatonovic-Kustrin, S. & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis* **22:** 717–727.

Anwar, M.R., Liu, D.L., Macadam, I. & Kelly, G. (2013). Adapting agriculture to climate change: a review. *Theoretical and Applied Climatology* **113:** 225–245.

Banerjee, S., Schlaeppi, K. & van der Heijden, M.G.A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology* **16:** 567–576.

Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris, D. & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors* **21:** 3758.

Bhat, K.A., Mir, R.A., Farooq, A., Manzoor, M., Hami, A., Allie, K.A., Wani, S.M., Khan, M.N., Sayyed, R.Z., Poczai, P., Almalki, W.H., Zargar, S.M. & Shah, A.A. (2022). Advances in nematode identification: a journey from fundamentals to evolutionary aspects. *Diversity* **14:** 536.

Bonneau, M., Gaba, S., Peyrard, N. & Sabbadin, R. (2014). Reinforcement learning-based design of sampling policies under cost constraints in Markov random fields: Application to weed map reconstruction. *Computational Statistics & Data Analysis* **72:** 30–44.

Breiman, L. (2001). Random Forests. *Machine Learning* **45:** 5 32.

Brito de Jesus, S., Vieira, D., Gheller, P., Cunha, B.P., Gallucci, F. & Fonseca, G. (2023). Machine learning algorithms accurately identify free-living marine nematode species. *PeerJ* **11:** e16216.

Chen, S.H. & Pollino, C.A. (2012). Good practice in Bayesian network modelling. *Environmental Modelling & Software* **37:** 134–145.

Cunningham, J.M., Koytiger, G., Sorger, P.K. & AlQuraishi, M. (2020). Biophysical prediction of protein-peptide interactions and signaling networks using machine learning. *Nature Methods* **17:** 175–183.

de Sousa, C.A.R. (2015). An overview on the Gaussian Fields and Harmonic Functions method for semi-supervised learning. *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. https://doi.org/10.1109/IJCNN.2015.7280491

Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y. & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture* **155:** 257–282.

Eli-Chukwu, N.C. (2019). Applications of artificial intelligence in agriculture: A review. *Engineering, Technology & Applied Science Research* **9:** 4377–4383.

Hoang, N.-D., Huynh, T.-C. & Tran, V.-D. (2021). Computer vision-based patched and unpatched pothole classification using machine learning approach optimized by forensic-based investigation metaheuristic. *Complexity* 2021: 3511375.

Jansson, J.K. & Hofmockel, K.S. (2020). Soil microbiomes and climate change. *Nature Reviews Microbiology* **18:** 35–46.

Jiang, J., Xing, F., Wang, C. & Zeng, X. (2018). Identification and analysis of rice yield-related candidate genes by walking on the functional network. *Frontiers in Plant Science* **9:** 1685.

Jones, J.T., Haegeman, A., Danchin, E.G.J., Gaur, H.S., Helder, J., Jones, M.G.K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J.E., Wesemael, W.M.L. & Perry, R.N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology* **14:** 946–961.

Kodinariya, T. & Makwana, P. (2013). Review on determining of cluster in K-means clustering. *International Journal of Advance Research in Computer Science and Management Studies* **1:** 90–95.

Kumar, V., Khan, M.R. & Walia, R.K. (2020). Crop loss estimations due to plant-parasitic nematodes in major crops in India. *National Academy Science Letters* **43:** 409–412.

Leroux, C., Jones, H., Clenet, A. & Tisseyre, B. (2019). Knowledge discovery and unsupervised detection of within-field yield defective observations. *Computers and Electronics in Agriculture* **156:** 645–659.

Ma, Y., Oliveira, R.S., Freitas, H. & Zhang, C. (2016). Biochemical and molecular mechanisms of plant-microbe-metal interactions: relevance for phytoremediation. *Frontiers in Plant Science* **7:** 918.

Mandal, H., Katel, S., Subedi, S., Shrestha, J., Acharya, N. & Acharya, P. (2021). Plant parasitic nematodes and their management in crop production: a review. *Journal of Agriculture and Natural Resources* **4:** 327–338.

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons* **4:** 51–62.

Ngo, M.T., Van Nguyen, M., Han, J.W., Kim, B., Kim, Y.K., Park, M.S., Kim, H. & Choi, G.J. (2021). Biocontrol potential of A*spergillus* species producing antimicrobial metabolites. *Frontiers in Microbiology* **12:** 804333.

Nowpada, R.S., Rao, K. & Siresha S. (2011). A new approach to evaluate characteristics involved in a fuzzy critical path method. *Advances in Fuzzy Mathematics* **6:** 217–226.

Pérez-Ortiz, M., Peña, J.M., Gutiérrez, P.A., Torres-Sánchez, J., Hervás-Martínez, C. & López-Granados, F. (2015). A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing* **37:** 533–544.

Pisner, D.A. & Schnyer, D.M. (2020). Support vector machine. In: Mechelli, A. & Vieira, S. (eds.), *Machine learning*. Academic Press, pp. 101–121.

Pun, T., Neupane, A. & Koech, R. (2023). A deep learning-based decision support tool for plant-parasitic nematode management. *Journal of Imaging* **9:** 240.

Qing, X., Peng, H., Ma, J., Zhang, Y.M., Li, H., Peng, D., Wang, X. & Long, T. (2022). Phylogeography of Chinese cereal cyst nematodes sheds lights on their origin and dispersal. *Evolutionary Applications* **15:** 1236–1248.

Ropelewska, E., Skwiercz, A. & Sobczak, M. (2023). Distinguishing cyst nematode species using image textures and artificial neural networks. *Agronomy* **13:** 2277.

Sert, E., Bar-Yam, Y. & Morales, A.J. (2020). Segregation dynamics with reinforcement learning and agent based modeling. *Scientific Reports* **10:** 11771.

Shabrina, N.H., Lika, R.A. & Indarti, S. (2023). Deep learning models for automatic identification of plant-parasitic nematode. *Artificial Intelligence in Agriculture* **7:** 1–12.

Sharma, S., Ang, J.B. & Fredriksson, P.G. (2021). Religiosity and climate change policies. *Energy Economics* **101:** 105414.

Shi, Y.-F., Yang, Z.-X., Ma, S., Kang, P.-L., Shang, C., Hu, P. & Liu, Z.-P. (2023). Machine learning for chemistry: basics and applications. *Engineering* **27:** 70–83.

Song, Y.Y. & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry* **27:** 130–135.

Sonnadara, D.U.J. & Jayewardene, D.R. (2015). A Markov chain probability model to describe wet and dry patterns of weather at Colombo. *Theoretical and Applied Climatology* **119:** 333–340.

Sun, B. & Park, B.B. (2017). Route choice modeling with support vector machine. *Transportation Research Procedia* **25:** 1806–1814.

Tang, Y. (2013). Deep Learning using linear support vector machines. arXiv preprint arXiv:1306.0239. 2013 Jun 2. https://doi.org/10.48550/arXiv.1306.0239

Xing, W. & Wang, B. (2017). Predictability and prediction of summer rainfall in the arid and semi-arid regions of China. *Climate Dynamics* **49:** 419–431.

Zhu, M., Meng, F. & Zhou, Y. (2013). Semisupervised clustering for networks based on fast affinity propagation. *Mathematical Problems in Engineering* 2013: 385265.

Zoppis, I., Mauri, G. & Dondi, R. (2018). Kernel Methods: Support vector machines. In: *Encyclopaedia of Bioinformatics and Computational Biology: ABC of Bioinformatics.* Vol. 1: Methods 2019, Elsevier, pp. 503–510.