Hindawi

Research Article

Prediction of Crop Yield by Support Vector Machine Coupled with Deep Learning Algorithm Procedures in Lower Kulfo Watershed of Ethiopia

Abebe Temesgen Ayalew [ID] **and Tarun Kumar Lohani** [ID]

Faculty of Hydraulic and Water Resources Engineering, Arba Minch University, Arba Minch, Ethiopia

Correspondence should be addressed to Tarun Kumar Lohani; tklohani@gmail.com

Sensible and judicious utilization of water for agriculture in conjunction with prediction techniques increases the crop yield. The Ethiopian economy relies on and is exclusively dependent on agricultural-based activities. Different soil compositions (nitrogen, phosphorous, and potassium), crop alternation, soil dampness, and climate conditions play an imperative contribution in cultivation. The primary purpose of this study was to conduct a machine learning approach which can be practiced dynamically for efficient farming at a low cost. The support vector machine (SVM) was applied as a machine learning procedure, whereas long short-term memory (LSTM) and the recurrent neural network (RNN) were considered as deep learning procedures. The research comprised a model that is combined with machine learning procedures (ANN, random forest, and decision tree) to know efficient and appropriate crop types. The planned model is improved through conducting deep learning methods incorporated to the existing practice for different crop condition. Pure data and related evidence are attained concerning the quantities of soil constituents desired through their expenditures distinctly. It delivers well precision as compared to the current model examining the specified documents and assisting the local agronomists in forecasting different types of crop and gain benefits. In RNN, LSTM, and SVM algorithms, the accuracy is determined as 96% which is comparatively preferable as compared to other machine learning procedures under different feature and crop types. The techniques are evaluated in terms of percentage in prediction accuracy. The results generated are important for agrarians, experts, researchers, and local farmers to maximize the crop productivity and help to enhance agriculture and climate change-related decisions, especially in low-to-middle-income countries.

1. Introduction

Agricultural farming in this present world has become more hi-tech in implementing various modern technologies to increase the agribusiness generating higher growth. Contemporary farming technologies have been recently adopted by developing countries that need a low capital investment but a higher return [1–4]. Agriculture is the backbone of the income source for developing countries such as Ethiopia. Numerous farming practices have been successfully implemented to increase the economy of the country but still a lot has to be done for sustainable growth of the country. 33.56% of the land has been used for conventional farming by an approximate population of around 120 million. A

traditional way of measuring crop yield has been practiced by almost all the farmers at the time of harvesting [5–7]. Farmers have been trained of recently developed procedures to increase their productivity level. On the contrary, a high rate of increase in fertilizers and indiscriminate use of pesticides have degraded the quality of soil and reduced the productivity. Expansion of harvesting has been alarming for farmers who intend to go for large and quick farming. Most of the farmers are unaware of the cons that affect their agricultural lands due to overuse of fertilizers [8–12]. The primary elements of the soil such as phosphorous, nitrogen, and potassium get distressed by unethical farming techniques [13–16]. In order to use computer-based farming, datasets are taken as static information to contemplate as

handling features [17–21]. Information from the field helps to know the data during initial and harvesting stages for numerous crop yields obtained from different platforms [22–26]. Artificial intelligence (AI) has been used as a planned arrangement determining multiple linear regression (MLR) towards identifying the model to accrue lots of information [12]; Sonal and Sandhya 2021. This study provides a general objective of integrating machine learning and deep learning algorithms for farm-level crop yield prediction in a smaller segment. Within this framework, three specific objectives were set: The first objective was to know the relation between the different parameters such as water, pesticides, area, and fertilizer usage with respect to the crop yield. The second objective was to evaluate the performance of the model. The third aim was to determine the level of accuracy for each algorithm and identify the best. Therefore, this study may provide the best suitable method to the agrarians and recommend numerous valuable yields. The finding of this research may help the local farmers for best adaptive farming for sustainable and long-term cultivation.

2. Study Area Description and Proposed Framework

2.1. Study Area Description. Kulfo watershed is one of the subwatersheds of Abaya-Chamo subbasin located between $37°28'00.94''$ and $37°34'03.34''$ east longitude and $6^0 03'49.19''$ and $6°03'11.71''$ north latitude. The watershed is the part of the Ethiopian Rift Valley in the southern region of the country. The Kulfo River is one of the prevailing rivers in the subbasin system which is originated from Guge Mountains, flowing towards east into Lake Chamo. It has a gauging station at Kulfo near the bridge. The average temperature lies between 23.05 C and 25.87 C. The maximum and minimum altitudes of the area are 3557 and 1192 m above mean sea level (a.m.s.l.), respectively (Figure 1).

2.2. Proposed Structure. The proposed structure for this research work driven from deep learning and machine learning procedures is executed to forecast suitable crop production from the agricultural land. An experiment-based approach is carried out in an entry crop data configuration by the provided model. The dominating crop type is selected by considering the present climate condition and soil composition including the respective parameters. An assimilated deep learning approach/algorithm is applied to realize abundant numerical platforms due to its applicability in achieving the optimal suitable crop for different alternative conditions. With the help of this platform, crops are projected precisely. The SVM procedure is accomplished with the three programming modes, namely, LSTM, RNN, and SVM which performed along with deep learning algorithms (Figure 2).

2.2.1. Details of the Programming Models (SVM, LSTM, and RNN). A support vector machine is a supervised machine learning technique mainly used for classification and regression problems [8, 11, 12]. Using statistical learning theory, the SVM algorithm was developed by Vapnik and Chervonenkis [12, 20]. It is about learning structure from given data and is capable of handling multiple continuous and categorical variables. The SVM model (Figure 3) represents different classes in a hyperplane in a multidimensional space. The aim of SVM is to categorize a dataset into different classes to find out the maximum marginal hyperplane (MMH).

2.2.2. LSTM (Long Short-Term Memory). An LSTM is a kind of RNN architecture. It is theoretically a more sophisticated recurrent neural network [14]. Instead of just having recurrence, it also has "gates" that regulate information flow through the unit. LSTMs were primarily initiated to solve the disappearing gradient problem of RNNs [14, 16]. They are often used over traditional, simple recurrent neural networks because they are also more computationally efficient (Figure 4).

2.2.3. RNN (Recurrent Neural Network). A recurrent neural network (RNN) is the type of artificial neural network (ANN) used in Apple's Siri and Google's voice search [11]. RNN remembers past inputs due to an internal memory which is quite useful for predicting stock prices, generating text, transcriptions, and machine translation. In the traditional neural network, the inputs and the outputs are independent of each other, whereas the output in RNN is dependent on prior elementals within the sequence (Figure 5). Recurrent networks also share parameters across each layer of the network [15]. In feedforward networks, there are different weights across each node, whereas RNN shares the same weights within each layer of the network. During gradient descent, the weights and basis are adjusted individually to reduce the loss.

2.2.4. Execution Steps

> Step 1: encoding input data comprising numerous set of parameters
>
> Step 2: encoding helpful packages and programmers
>
> Step 3: preprocessing of data manipulation
>
> Step 4: data division required to make data entry format into preparation set as well as examining set
>
> Step 5: the model assembled using machine learning through the support vector platform (SVM) and deep learning algorithm procedures
>
> Step 6: in the examining stage, the efficiency of the model is checked, and when there is an error in the input data, there will be false forecasting to be executed

2.3. Detail Description of Data Entry. The data comprising different agricultural parameters are considered to run the model. The remaining data were taken as feature entry format. The proposed size of the data entry is around 7700 kb. Numerous parameters used in this data entry
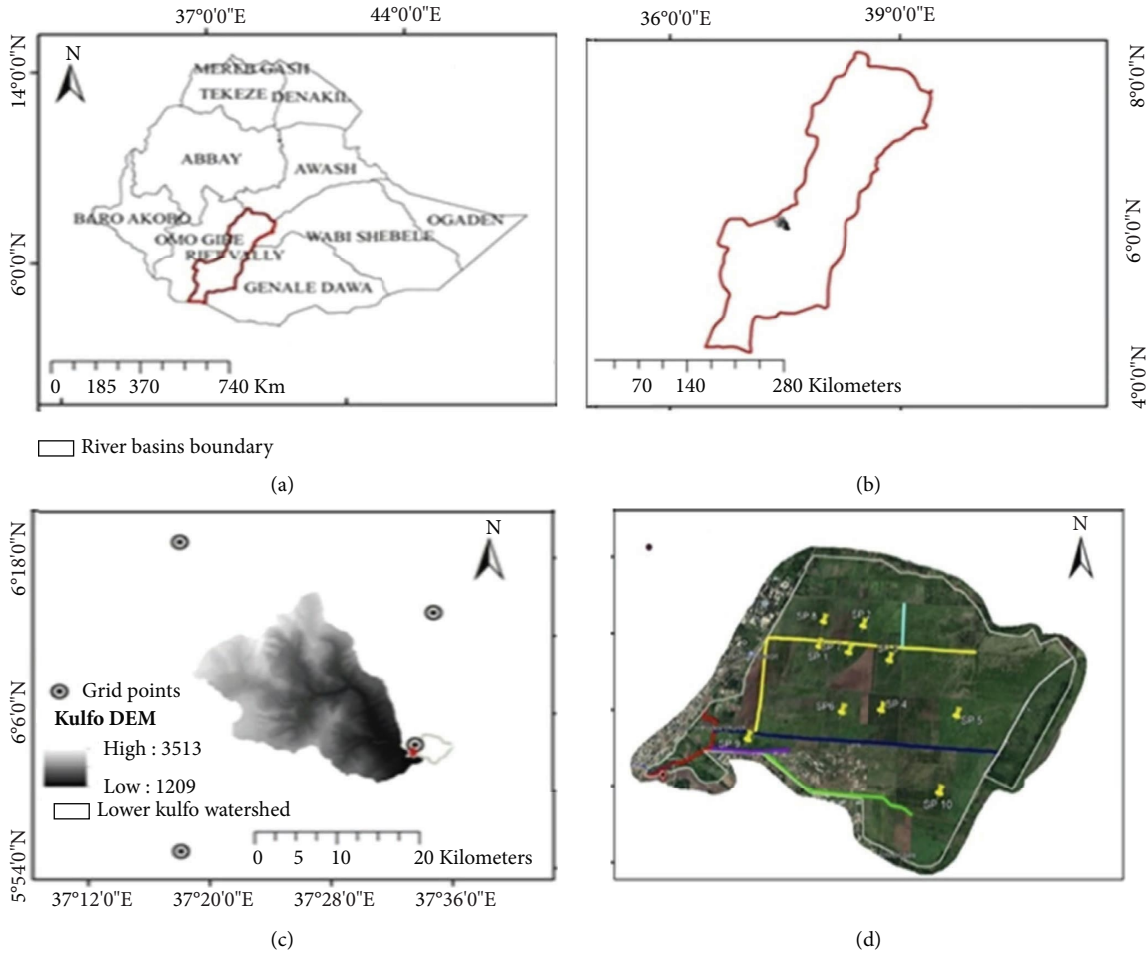
Figure 1: Study area: (a) Ethiopian major river basins, (b) Rift valley basin, (c) Kulfo watershed, and (d) lower Kulfo watershed.
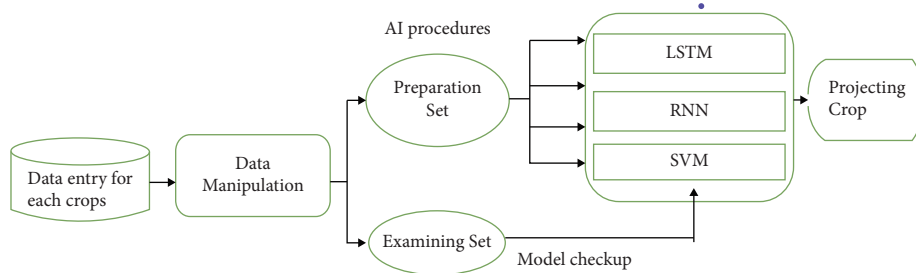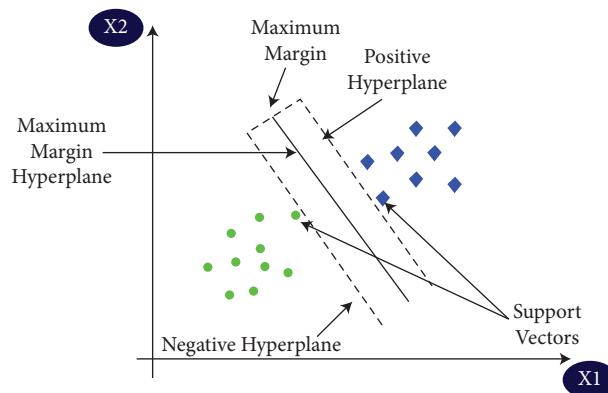

Figure 2: Model setup for projecting crops.
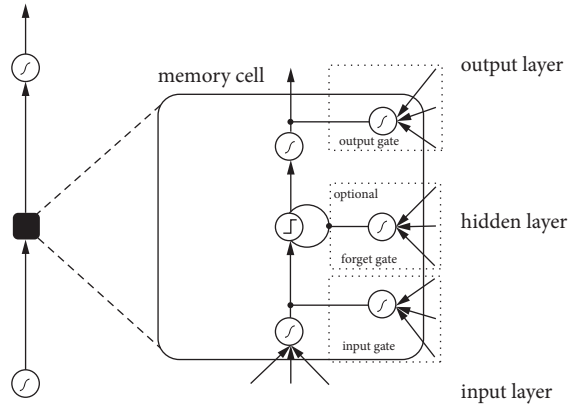

Figure 3: Simple SVM machine [16].

Figure 4: Simple LSTM network [12].

Figure 5: Simple RNN network [11].
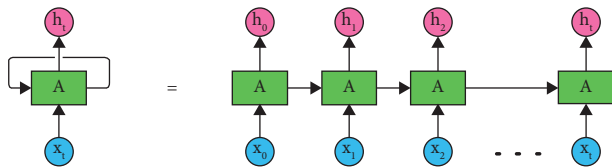
Table 1: Crop dataset and its arrangement.

Crop	pH	Temperature (°C)	Rainfall (mm)	Relative humidity
			0.0	54.0
			0.0	53.0
			0.0	54.0
			0.0	68.0
			0.0	68.0
			0.0	66.0
			0.0	58.0
			0.0	58.0
			0.0	59.0
			0.0	63.0
			3.7	66.0
Maize	6.5	24	0.0	52.0
			7.6	45.0
			20.0	43.0
			0.0	48.0
			0.0	39.0
			0.0	47.0
			2.0	54.0
			0.0	55.0
			8.0	46.0
			1.4	45.0
			2.7	45.0
			0.5	49.0
			0.0	51.0

comprise precipitation, relative humidity, temperature, pH, and area. The dominant crops used for these data include banana, maize, wheat, potato, and tomato. Different values are available with numerous forecasting parameters for a single set of crop. For example, while considering the crop data entry as banana, several datasets are considered for the forecast parameters as compared to a data entry format of values existing within the dataset. The procedure is similar for the rest crops present in the data entry format.

Table 1 shows a small portion of the crop dataset. The major crop type is maize, with the corresponding pH value of the soil sample being 6.5 and its temperature being 24.

2.4. Algorithms Used

(i) The support vector machine (SVM) platform

Step 1: mandatory package is imported
Step 2: uploading the input entry values
Step 3: the required quantity of arrangement from the total data entry is identified
Step 4: the SVM limits are plotted to their original or reference data
Step 5: consistent parameter values are defined
Step 6: an SVM classifier is produced

(ii) Long short-term memory (LSTM)

(iii) Recurrent neural network (RNN)

3. Evaluation of the Method Section

The performance of the planned model is calibrated in order to get the exact result. Numerous formulas are adopted to get the accuracy of the final result. Some of the formulas are listed as equations (1) and (2) [6, 7]:

$$\text{Accuracy} = \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{FP} + \text{TP} + \text{FN})}, \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})},$$
$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \quad (2)$$

Note: TP: when the cases are predicted as +ve and are +ve in actual. FP: when the cases are predicted as as +ve and are -ve in actual. TN: when the cases are predicted as -ve and are -ve in actual. FN: when the cases are predicted as -ve and are +ve in actual.

3.1. Implementation and Result. Loading different datasets that already exist will be alarming to start at the implementation stage. This is followed by necessary libraries and packages imported to proceed and perform preprocessing raw data. Raw data are classified as test and trained data. Moreover, a model is made, and AI algorithms are developed. This will help to know the best and appropriate crop that could be grown in a specific farm land.

3.2. Result Analysis. This machine learning-based crop yield prediction is applied through a crop raw data entry which has already been compiled from various sources from the information on the predictive model which contains the test set and the trained set. The trained set is collected from the historical survey data of the farm land, and the test set is collected simply from simple survey data. The response function after the model is run provides a graphical
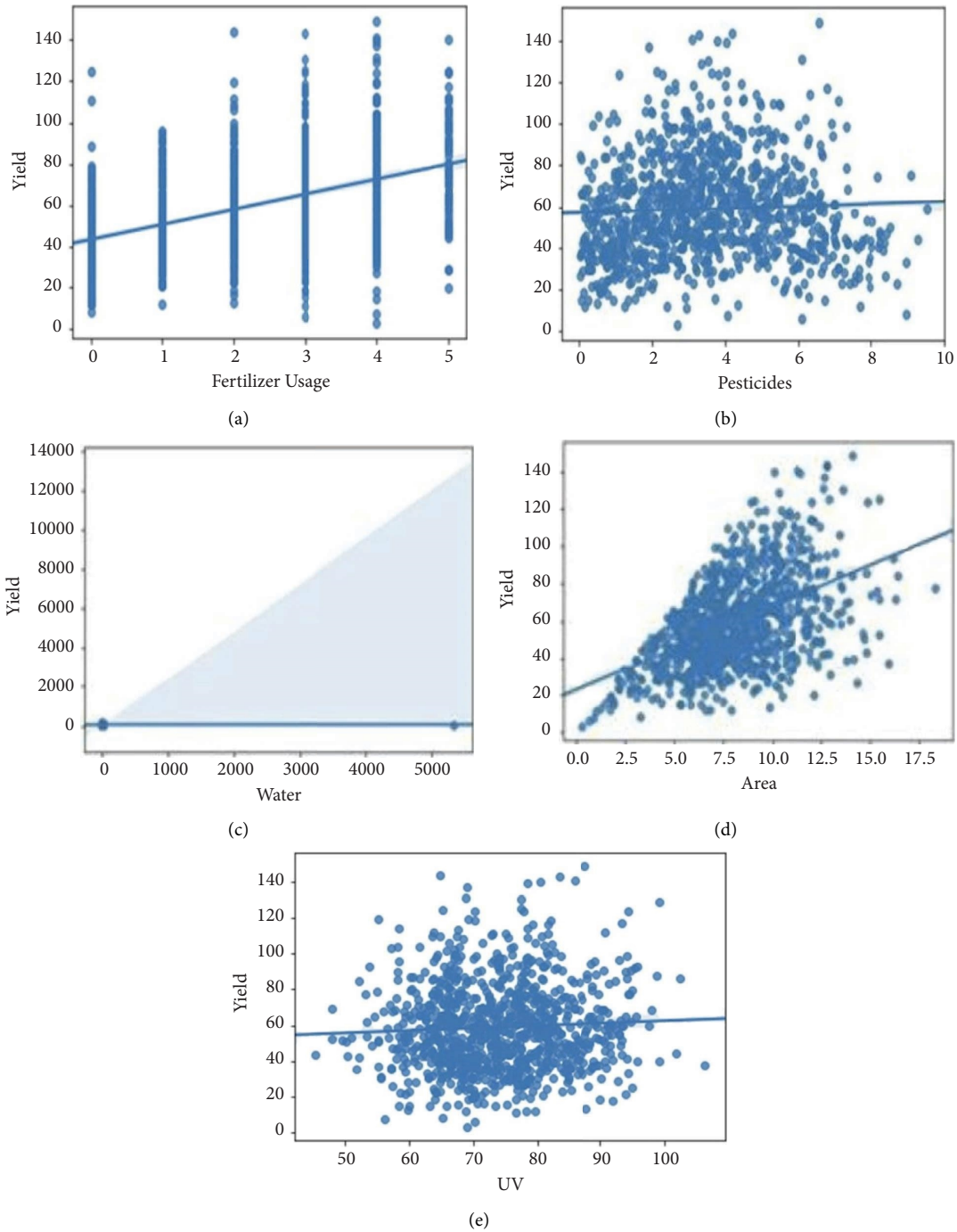
(a)

(b)

(c)

(d)

(e)

Figure 6: (a) Yield vs. fertilizer usage, (b) yield vs. pesticides, (c) yield vs. water, (d) yield vs. area, and (e) yield vs. uv.

presentation (Figures 6(a)–6(e)) for the parameters mentioned in the figures, namely, water, pesticides, area, uv, and fertilizer usage. Finally, the total yield of the crop is determined with respect to the relation between those constraints.

Those figures show the relation between the different parameters such as water, pesticides, area, uv, and fertilizer usage with respect to the crop yield. The support vector machine platform helps to configure the graphs. A comparative analysis has been presented regarding yield with respect to the fertilizer use (Figure 6(a)). Besides, in Figure 6(b), the yield is estimated in contradiction of pesticides. The yield of the crop is determined through examining water as mentioned (Figure 6(c)). Figure 6(d) also represents interdependent relation between the yield and area. The yield is estimated in contradiction of uv in Figure 6(e). From all the determinate factors in the plot, each value is taken from each feature dataset or arrangement, and then accordingly, the crop yield is determined.

TABLE 2: Evaluation indicators.

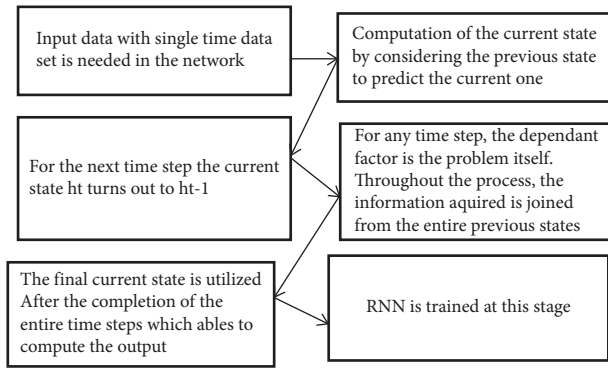| Machine learning procedures	Features	Crops	Accuracy (%)
DT, artificial neural network, and RF for existing condition	Temperature, precipitation, position, and soil	Banana, maize, and wheat	92
LSTM, RNN, and SVM (for the proposed activities)	Temperature, precipitation, pH value, relative humidity, and area	Wheat, tomato, potato, maize, and banana	96

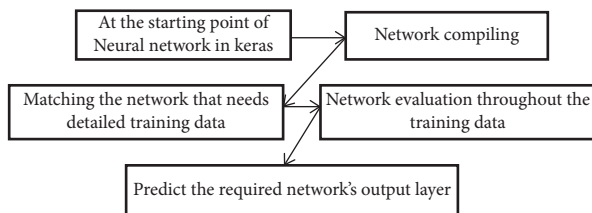Figure 7: Steps throughout the recurrent neural network.

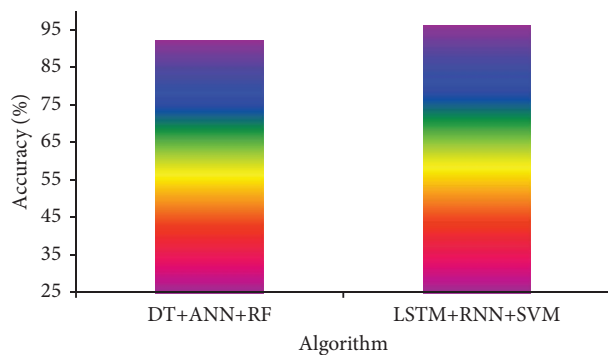Figure 8: Various steps in long short-term memory.

Figure 9: Performance graph displaying accuracy level.

3.3. Evaluation of the Performance of the Model. Several parameters are used for analysis considering the dataset. Those constrained parameters are temperature, precipitation, location of soil, relative humidity, and area. The dataset is examined, and the better yield is predicted. There are also certain crops which are taken for yielding on a specific portion of land. These crops include wheat, tomato, potato maize, and banana.

Applying different machine learning algorithms under different features and crop types, the accuracy is analysed (Table 2). The accuracy level is 92% from those algorithms including random forest algorithms and artificial neural networks (ANNs). Besides, for the recurrent neural network (RNN) (Figure 7), long short-term memory (LSTM) (Figure 8), and support vector machine (SVM) algorithms [27, 28], the accuracy is determined as 96% and the less accuracy around 86.3% as compared to this study [29]. This study has its own importance since it gives high accuracy with the use of integrated machine learning algorithms.

Deep learning algorithms along with machine learning procedures play a great role in predicting crop yield with better accuracy as compared to other machine learning procedures (Table 2).

Figure 9 shows the accuracy assessment for both the algorithms. The approach bar in the first plot has presented the accuracy of 92% for the ANN supported by the random forest algorithms. The approach bar in the second plot has shown the accuracy of 96% when applying LSTM, RNN, and SVM algorithms together. Hence, the approach bar in the second plot gives a better accuracy.

4. Conclusion

The planned model built through AI procedures to minimize the difficulties faced by agrarians, shortage of information, or awareness and equipped skill of farming in diverse weather and soil conditions are well articulated. The model is produced by machine learning (SVM) and deep learning methods. For modification or to improve the current cropping status, this platform provides a remarkable result. Moreover, it is easy to decide that there is an improvement in the precision of the research work as compared to the current study that used other procedures for forecasting and predicting of crops. The total accuracy is estimated as 96%. Numerous approaches may be developed in the future to portray interface, exhibiting flexible and versatile applications. Almost all local farmers want an automated system to get sufficient information on which crop type provides best yield. With this, even in the absence of the farmers, the work can be managed at that specific time, without facing any kind of difficulty or loss. Modern farming techniques will be extensively adapted and in turn will help the farmers to increase productivity in every season of farming. Comprehensive machine and deep learning approaches are adopted to increase the crop yield from the output of deep learning algorithms (SVM, LSTM, and RNN). The finding of this research provides a platform to the local farmers for best farming techniques for sustainable and long-term cultivation. The current research can be extended into performing further analysis and forecasting the factors that influence crop yield. A larger dataset and more historically accurate data about the environment and weather during each crop year are required to identify the best-performing model between deep learning and machine learning models. More deep learning models need to be tested on the dataset. In the field of crop yield prediction, remote sensing data could be merged with the localised statistical data to improve the model's performance.

Data Availability

The data generated or analysed during the study are available upon reasonable request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Abebe Temesgen Ayalew collected the data, designed the work, and analysed and interpreted the data for the work. Tarun Kumar Lohani drew the figures and graphs, drafted the whole manuscript, and reviewed it critically. Both the authors have read the final draft thoroughly and approved for publishing.

Acknowledgments

References

[1] F. Perez, "Jupyter notebook (version 6.1.5): project jupyter," 2014, https://jupyter.org/index.html.

[2] Y. X. Su, H. Xu, and L. J. Yan, "Support vector machine-based open crop model (SBOCM): case of rice production in China," *Saudi Journal of Biological Sciences*, vol. 24, no. 3, pp. 537–547, 2017.

[3] O. E. Apolo-Apolo, J. Martínez-Guanter, G. Egea, P. Raja, and M. Pérez-Ruiz, "Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV," *European Journal of Agronomy*, vol. 115, Article ID 126030, 2020.

[4] B. Peng, K. Guan, W. Zhou et al., "Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop yield prediction," *International Journal of Applied Earth Observation and Geoinformation*, vol. 90, Article ID 102126, 2020.

[5] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review," *Computers and Electronics in Agriculture*, vol. 151, pp. 61–69, 2018.

[6] M. K. Dharani, R. Thamilselvan, P. Natesan, P. C. D. Kalaivaani, and S. Santhoshkumar, "Review on crop prediction using deep learning techniques," in *Journal of Physics: Conference Series*vol. 1767, IOP Publishing, Article ID 012026, 2021.

[7] H. R. Bedane, K. T. Beketie, E. E. Fantahun, G. L. Feyisa, and F. A. Anose, "The impact of rainfall variability and crop production on vertisols in the central highlands of Ethiopia," *Environmental Systems Research*, vol. 11, no. 1, p. 26, 2022.

[8] D. B. Lobell, K. N. Cahill, and C. B. Field, "Historical effects of temperature and precipitation on California crop yields," *Climatic Change*, vol. 81, no. 2, pp. 187–203, 2007.

[9] P. Filippi, E. J. Jones, N. S. Wimalathunge et al., "An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning," *Precision Agriculture*, vol. 20, no. 5, pp. 1015–1029, 2019.

[10] Food and Agriculture Organization of the United Nations, "Crops. FAO FAOSTAT," 2019, http://www.fao.org/faostat/en/#data/QC.

[11] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning–method overview and review of use for fruit detection and yield estimation," *Computers and Electronics in Agriculture*, vol. 162, pp. 219–234, 2019.

[12] Q. Zhang, Y. Liu, C. Gong, Y. Chen, and H. Yu, "Applications of deep learning for dense scenes analysis in agriculture: a review," *Sensors*, vol. 20, no. 5, p. 1520, 2020.

[13] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers in Plant Science*, vol. 10, p. 621, 2019.

[14] S. Lee, Y. Jeong, S. Son, and B. Lee, "A self-predictable crop yield platform (SCYP) based on crop diseases using deep learning," *Sustainability*, vol. 11, no. 13, p. 3637, 2019.

[15] H. Jiang, H. Hu, R. Zhong et al., "A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: a case study of the US Corn Belt at the county level," *Global Change Biology*, vol. 26, no. 3, pp. 1754–1766, 2020.

[16] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritschi, "Soybean yield prediction from UAV using multimodal data fusion and deep learning," *Remote Sensing of Environment*, vol. 237, Article ID 111599, 2020.

[17] P. W. T. Oliphant, "Anaconda (version 4.9.2): anaconda inc," 2012, https://www.anaconda.com/.

[18] N. Hani, P. Roy, and V. Isler, "A comparative study of fruit detection and counting methods for yield mapping in apple orchards," *Journal of Field Robotics*, vol. 37, no. 2, pp. 263–282, 2020.

[19] S. Khaki, L. Wang, and S. V. Archontoulis, "A cnn-rnn framework for crop yield prediction," *Frontiers in Plant Science*, vol. 10, p. 1750, 2019.

[20] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: a systematic literature review," *Computers and Electronics in Agriculture*, vol. 177, Article ID 105709, 2020.

[21] S. Agarwal and S. Tarar, "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms," in *Journal of Physics: Conference Series*vol. 1714, IOP Publishing, Article ID 012012, 2021.

[22] R. Gv, "Python language reference (version 3.7.9): research center centrum wiskunde & informatica (CWI)," 1989, https://www.python.org/.

[23] N. Kim and Y. W. Lee, "Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 34, no. 4, pp. 383–390, 2016.

[24] X. Xu, P. Gao, X. Zhu et al., "Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China," *Ecological Indicators*, vol. 101, pp. 943–953, 2019.

[25] S. Fountas, N. Mylonas, I. Malounas, E. Rodias, C. Hellmannsantos, and E. Pekkeriet, "Agricultural robotics for field operations," *Sensors*, vol. 20, no. 9, p. 2672, 2020.

[26] N. Chergui and M. Kechadi, "Data analytics for crop management: a big data view," *Journal of Big Data*, vol. 9, no. 1, p. 123, 2022.

[27] Z. G. Zhou, B. Chen, Z. Li, and C. Li, "The use of LSTM-based RNN and SVM models to detect ludian coseismic landslides in time series images," in *Journal of Physics: Conference Series*vol. 1631, IOP Publishing, Article ID 012085, 2020.

[28] A. Satan, A. Khamzina, D. Toktarbayev, Z. Sotsial, I. Bapiyev, and N. Zhakiyev, "Comparative LSTM and SVM machine learning approaches for energy consumption prediction: case study in akmola," in *Proceedings of the 2022 International Conference on Smart Information Systems and Technologies (SIST)*, pp. 1–7, IEEE, Nur-Sultan, Kazakhstan, April 2022.

[29] S. Iniyan, V. Akhil Varma, and C. Teja Naidu, "Crop yield prediction using machine learning techniques," *Advances in Engineering Software*, vol. 175, Article ID 103326, 2023.