

Communication

# Hybrid CNN-SVM Classifier Approaches to Process Semi-Structured Data in Sugarcane Yield Forecasting Production

Debnath Bhattacharyya <sup>1</sup>, Eali Stephen Neal Joshua <sup>2</sup>, N. Thirupathi Rao <sup>3</sup> and Tai-hoon Kim <sup>4,\*</sup>

<sup>1</sup> Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur 522302, Andhra-Pradesh, India

<sup>2</sup> Department of Computer Science & Engineering, Gitam (Deemed to be University), Gandhi Nagar, Rushikonda, Visakhapatnam 530045, Andhra-Pradesh, India

<sup>3</sup> Department of Computer Science & Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam 530049, Andhra-Pradesh, India

<sup>4</sup> School of Electrical and Computer Engineering, Yeosu Campus, Chonnam National University, 50, Daehak-ro, Yeosu-si 59626, Jeollanam-do, Republic of Korea

\* Correspondence: taihoonn@chonnam.ac.kr; Tel.: +82-10-8592-4900

**Abstract:** Information communication technology (ICT) breakthroughs have boosted global social and economic progress. Most rural Indians rely on agriculture for income. The growing population requires modern agricultural practices. ICT is crucial for educating farmers on how to be environmentally friendly. It helps them create more food by solving a variety of challenges. India's sugarcane crop is popular and lucrative. Long-term crops that require water do not need specific soil. They need water; the ground should always have adequate water due to the link between cane growth and evaporation. This research focuses on forecasting soil moisture and classifying sugarcane output; sugarcane has so many applications that it must be categorized. This research examines these claims: The first phase model predicts soil moisture using two-level ensemble classifiers. Secondly, to boost performance, the proposed ensemble model integrates the Gaussian probabilistic method (GPM), the convolutional neural network (CNN), and support vector machines (SVM). The suggested approach aims to correctly anticipate future soil moisture measurements affecting crop growth and cultivation. The proposed model is 89.53% more accurate than conventional neural network classifiers. The recommended models' outcomes will assist farmers and agricultural authorities in boosting production.

**Keywords:** Information communication technology; agriculture; ensemble learning; Gaussian probabilistic method function; convolutional neural network; support vector machines



**Citation:** Bhattacharyya, D.; Joshua, E.S.N.; Rao, N.T.; Kim, T.-h. Hybrid CNN-SVM Classifier Approaches to Process Semi-Structured Data in Sugarcane Yield Forecasting Production. *Agronomy* **2023**, *13*, 1169. <https://doi.org/10.3390/agronomy13041169>

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 4 March 2023

Revised: 10 April 2023

Accepted: 12 April 2023

Published: 20 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Precision agriculture uses cutting-edge technology to improve crop yield and quality while meeting growing food demand [1]. Farmers may optimize the efficacy of their seeds, water, fertilizers, and other inputs by employing advanced farming techniques. The goal is to maximize financial benefit while reducing expenditures and the effects on the natural environment by using the correct parameters at the right time and place.

Farmers must monitor soil moisture. This entails obtaining a more precise image of the soil's needs and available water. An amount of 70% of the world's water is utilized for agriculture, 20% is utilized for industry, and 10% is utilized for residences, according to UN figures (<http://www.worldometers.info/>) (accessed on 10 February 2022) [2]. To monitor and plan irrigation, you must know the soil's water content. Sugarcane is a popular crop in India. It is the critical raw material for the nation's second-largest agricultural sector, behind only the textile industry.

Additionally, it serves as the foundation for producing every significant sweetener in the country. Long-term growth does not need a specific soil type, but it requires much

water. The soil must have enough water during the growing season, and cane growth depends on how much cane sheds. In addition to categorizing sugarcane yield, the research aims to anticipate soil moisture.

#### *Hybrid Model for Better Sugarcane Yield Production*

Agriculture has since incorporated soft computing to manage crops, anticipate agricultural yields, and monitor environmental conditions [3]. In India, sugarcane is one of the most important crops [4].

Sugarcane businesses provide significant benefits to rural areas in terms of the extra cash they generate and the job possibilities they make available. The proposed study will investigate how different soft computing technologies may be used to improve the categorization and forecasting of agricultural data to find better solutions to existing problems.

This section offers a condensed summary of the critical issues and contributions that this study has made.

- In order to provide predictions about the amount of moisture present in the soil, a mixed model was used. In order to do this, the most advantageous aspects of the GPM, CNN, and SVM algorithms may be merged.
- CNN is used to sort the data on the production of sugarcane.

According to a thorough literature survey, CNN analysis is helpful for crop management and yield research in precision agriculture. They improve over time by uncovering new patterns, not in the training datasets.

Ensemble learning's power comes from its capacity to train several algorithms and combine their results to obtain more accurate conclusions. The hybrid model, an ensemble technique, produces accurate predictions. To make hybrid solutions for real-world problems, you must combine at least two methods to compensate for each other's flaws. Because agriculture generates so much data, precision agriculture applications require various data processing technologies.

Academics have created algorithms for agricultural data applications. These algorithms will help farmers and agronomists manage crops and adopt agricultural strategies to guarantee a continual crop supply. This study uses neural networks, GPM, and SVM algorithms to improve agricultural data prediction and categorization.

The primary objective of this research is to enhance the precision of agricultural yield in sugarcane production in agriculture.

- Sugarcane production has been the most prominent crop in Andhra Pradesh, India.
- Certain soil moisture methods may be deployed for better crop yield.
- Developing forecasting models for soil moisture data sets using convolutional neural network ensembles was performed. To do this, the attributes of SVM and Gaussian function were combined. Hybrid model (CNN + SVM + Gaussian) approaches are being used to train the model's parameters to increase the accuracy of a model that detects sugarcane yields.

## **2. Related Works**

Precision agriculture and smart farming need rapid disease diagnosis [5]. Farmers may benefit from using plant pictures and existing machine-learning algorithms to detect crop illnesses early. Modern diagnostics may help rural farmers save their harvests. Yet, current technologies cannot accurately identify diseases. Academics and software developers are interested in how to pick a machine-learning model and scale it down to run efficiently on a device at the network's perimeter. This study yields a model for the Plant-Village and Plant-Doc datasets [6]. The technique relies on deep learning model performance assessment research. The analysis found that the proposed model is similar to the most sophisticated option on the market. Quantization ensures that the reduction in model size to suit an edge device will not affect system performance.

The N nutrition index helps farmers determine whether their crops have enough nitrogen (N) for safe harvesting. Nitrogen utilization efficiency (NUE) [7] will rise, reducing environmental damage. However, active canopy sensor data cannot accurately detect the crop's N condition or topdressing N quantity. Reflectance data from several active crop sensors is unpredictable because many environmental and management elements vary by location. To ensure food security and long-term growth, further study is needed to develop ways for monitoring and prescribing N levels throughout the growing season in various contexts.

The average phosphorus surplus in the Andhra-Pradesh and Telangana regions was 0.8 kg per acre per year. However, this varied by location. When inputs were lowered in regions with big phosphorus surpluses and easy-to-access soil, the region's yearly phosphorus budget revealed plenty of potential for improvement [8,9].

Agriculture is important to India's economy [10]. Most farmers know little about soil preservation. Grow the right crops and use the right fertilizers to maximise land production. Modern agriculture depends on technology. More productive and efficient agriculture practices are replacing older ones. A rising number of individuals are concerned about fertilizer consumption. Farmers may use several machine learning algorithms to determine fertilizer quantities. Fertilizer consumption rates vary according to crop-specific fertilizer needs. Machine learning (ML) researchers aim to create computer programs that can explore massive amounts of data for patterns and use those patterns to anticipate fresh data. Machine learning (ML) methods assist in solving prediction and classification issues. Machine learning, utilized in many agricultural contexts, helps farmers overcome their biggest challenges. This experiment tested the support vector machine (SVM), decision tree (DT), and CNN for soil nutritional analysis. The CNN algorithm is 4% more accurate than other algorithms [11,12].

New technologies affect farming, textiles, vehicles, engineering, and education. AI has shaped educational practices throughout time. This study [13] examines how current AI advances may influence new educational methods. The research created a useful random forest algorithm teaching methods learning companion. The proposed random forest technique performed inconsistently during training, but, thereafter, it performed better. The recommended strategy outperformed the support vector machine, a hesitant fuzzy set, and the overall expert score in the 30–35 range of the final weighted evaluation. 7%, 6%, and 17% were the results of this strategy.

Villupuram is a rural Tamil Nadu city. This research will focus on future climate and agriculture harvest advice [14]. The seasonality of agricultural production is another aspect that must be considered. This work computed 1990 agricultural production using Villupuram district data from 1990–1991 and 2019–2020. Moreover, agricultural production patterns will be examined. IMD forecasts weather and environment. Seasons alter the maximum temperature and lowest rainfall. SVM is the most reliable machine-learning method for seasonal output forecasting. We will also evaluate the SVM machine algorithm's efficiency and accuracy. A support vector machine (SVM) [15] showed that the seasonal temperature scale will accurately predict the global average temperature in 20 years. Researchers used agronomic practices to show when and how to grow different crops throughout this research.

Humans are endangering Madagascar's unique vegetation and fauna. This article presents data and analysis on the most prevalent and vital threats, collections, and programs, as well as protected area extent and breadth, to evaluate Madagascar's terrestrial and freshwater biodiversity's conservation status. An amount of 10.4% of Madagascar is preserved as terrestrial parks and reserves. Ninety-seven percent of freshwater fish, amphibians, reptiles, birds, and mammals, and 67% of native plants, live inside its boundaries. Rarer species exist (at least 79.6 percent of endangered plant species and 97.3 percent of endangered animal species reside in at least one protected location). Overexploitation of biological resources and unsustainable farming are the most significant risks to biodiversity, according to Bayesian neural network plant research and International Union

for Conservation of Nature (IUCN) Red List evaluations. We concentrate on five action levels to ensure that conservation and ecological restoration goals, methods, and activities address complex underlying and interrelated causes and benefit Madagascar's biodiversity and people. Due to cassava's weekly price increases, it is hard to anticipate agricultural expenditures and similar group inputs. Machine learning may be used to calculate cassava costs in the future [16,17].

The Office of Agricultural Economics under the Ministry of Agriculture and Cooperatives studied cassava prices from January 2005 to February 2022, which represents strategy development using the data. The support vector machine approach, incorporating potato and garlic prices, had the lowest root mean squared error (RMSE) [18] of 0.10 compared to the previous methodology that utilized the same information. Comparing the support vector machine to the previous approach revealed this. The recommended technique is 0.61% more successful, despite its 3.45% MAPE. The penultimate phase of price clustering, K-mean analysis, gave a December peak pricing period of 14.08%. The study would help the agricultural firm maximize profits by guiding its planting strategies.

It is essential when formulating food policy and ensuring there is enough food for everyone, to accurately predict what the winter wheat yields will be in each region ahead of time. This is especially important when considering the growing population and the changing environment. Artificial intelligence (AI) and large amounts of data in agriculture are the foundations of intelligent agriculture, which aims to help the agricultural industry find effective and cheap solutions to the challenges it faces [19]. Profound learning-based crop production predictions have emerged in recent years as an essential tool for controlling agricultural operations' output. In this study, we offer a Bayesian optimization-based strategy for predicting winter wheat production by fusing data from several sources using a long-term and short-term memory model. Our method can provide more accurate results than previous methods (BO-LSTM) [20]. Using multi-source data as input variables, analysis of BO-LSTM, support vector machine (SVM), and most minor absolute shrinkage and selection operator (Lasso) give the capability to make predictions. The results showed that Bayesian optimization might be used to adjust deep learning hyperparameters successfully.

The recent advancements in information and communication technology have impacted many aspects of the global economy. The rise of artificial intelligence (AI) and data science, in addition to the ubiquitous availability of digital technology, are likely to be credited as the primary drivers behind the arrival of digital agriculture. Farmers now have access to new farming practices that may increase productivity without affecting the ecological health of the land they work on as a residue of total agricultural improvement. The most recent developments in digital technology and data science have made it possible for agronomists, farmers, and other experts to acquire and analyze large amounts of agricultural data. This allows them to gain insight into farming operations and make better decisions. [21] examines how data mining methods may be used in contemporary agricultural technologies. [22] offers a high-level overview of agricultural yield management, focusing on crop production and monitoring as the two most important aspects to discuss. When the data mining approaches for monitoring agricultural production have been sorted into their basic categories, several studies that focus on analytics are discussed and evaluated. After that, a comprehensive analysis and discussion of big data's influence on agriculture follows.

The algorithm for predicting time series in decision support systems has seen substantial advancements in recent years. Automating and improving prediction models by using machine learning (ML) methods to analyze time series may be possible. This would come about when managing enormous amounts of data becomes more accessible. The most important thing this research brings is a hybrid model that performs better than previous forecasting models. To do this, its innovative take on the traditional Kalman filter is performed by incorporating machine learning techniques, such as support vector regression (SVR) [23] and nonlinear autoregressive (NAR) [24] neural networks into the mix. The proposed hybrid model addresses the problem of high error fluctuation in time

series data convergence by combining an updated Kalman filter strategy with a machine learning technique as a correction factor to estimate model error. The ML technique serves as a correction factor to estimate model error. The findings indicate that, compared to traditional and alternative Kalman filter models [25], our hybrid models have a goodness of fit of more than 0.95 and much lower root mean square and absolute mean errors. The method's applicability to various settings was shown further in two different ways.

### *Gap Analysis*

The thorough literature review [5–25] indicates the need to improve crop yield production. The authors offer a heavyweight model that calls for extended training and generates results that vary, depending on the computer. No matter how complicated the machine learning model is, domain experts cannot understand how it works. They have a hard time maintaining their consistency. Recent research has shown that all models, particularly machine learning models, can be accurate. However, it is pointless to utilize them if the pitching expert disagrees with the model. To clarify the decision-making process, to increase farmers' trust, and to normalize attitudes towards the concept, all of which would increase the likelihood of its eventual adoption, the goal is to use the lightweight process model and methods, such as integrating the SVM, as well as the radial basis function, with the ensemble learning techniques. This will be accomplished by using the proposed classifier.

The soil's moisture content is determined by its current water content. This factor affects air temperature, precipitation, and other weather aspects. Agriculture, precision farming, and drought monitoring benefit from their usage. When deciding when and how much to irrigate, growers must monitor soil moisture [26]. This increases output, improves product quality, and reduces disease and irrigation costs. Most Indians make their living via agriculture. Population growth will continue as a signifier for water conservation for human consumption, with the rest used for agricultural irrigation. You must understand the crop's water needs to meet this growing need. In locations with little water, monitoring soil moisture for effective irrigation management is increasingly important. Soil moisture explains how water and energy are moved from the ground surface to the atmosphere.

Real-time soil moisture data is needed to predict the weather, monitor floods, identify droughts early, and calculate crop growth. Forecasting is constructing a model to predict the future based on current knowledge. The capacity to accurately anticipate soil moisture affects crop growth, cultivation, and soil temperature. The challenge is creating a system that can continuously evaluate soil moisture levels. Researchers studied several statistical techniques, algorithms, and tools to construct ensemble models for weather forecasting and assessing soil moisture. The created ensembles' performance is estimated using a statistical method. A neural network-based annealing method was used to anticipate soil moisture. The suggested technique outperforms BPN and GPM networks [27]. A network-based model that can predict tomorrow's ground temperature. The model received ground and air temperatures. A good prediction helps navigate the unpredictable future. Random forest, bagging, boosting, voting, stacking, and boosting are popular data-gathering methods.

To precisely measure moisture in the soil, we must also have excellent and reliable sensors, such as quartz sensors that work on the principle of open capacitive sensitive elements pushed to different depths into the ground, such as those described in ref: [28,29].

## **3. Materials and Methods**

### *3.1. The Proposed Ensemble Model for Soil Moisture (Hybrid Model)*

#### **Introduction to the Model Deployment**

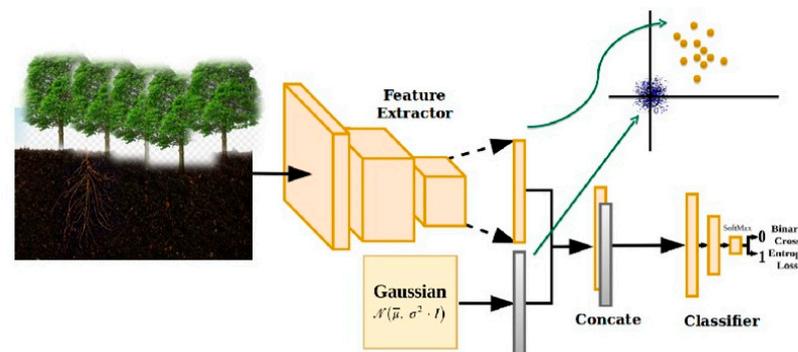
A network of wireless soil moisture sensors is placed to assess how wet the top layer of soil is. The base station stored the sensor's time series data. The soil moisture data from the base station are analyzed to fix the problem.

The study presented aims to enhance soil moisture forecasting. The suggested ensemble model comprises GPM, CNN, and SVM individual learners. The ZeroR classifier

provides predictions and is a meta-learner. The ZeroR classifier is the easiest to use and can predict the most frequent category. It helps to compare classifiers to a standard.

### 3.2. Gaussian Probabilistic Method for CNN-SVM

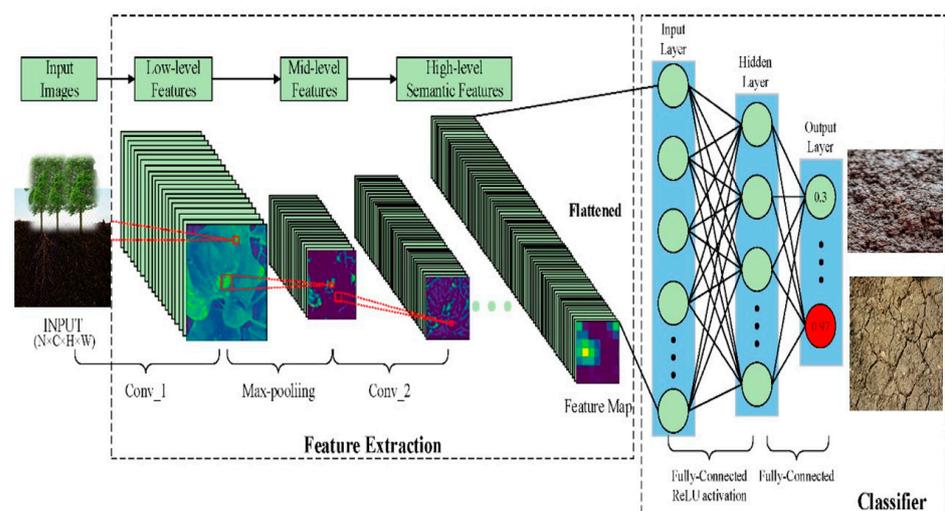
The Gaussian probabilistic functions were used with the CNN. The network's output is a linear blend of the inputs' GPM functions [30] and neuronal characteristics. Because it uses supervised learning, it can represent nonlinear data and help find the proper application. Input, concealed, and output are GPM network layers (Figure 1). GPM allows faster learning, better approximation, and more accessible network design than other CNNs. We will utilize least-squares mean-squared to determine optimum weights.



**Figure 1.** Gaussian probabilistic for the Soil Perception for adding nonlinearity.

### 3.3. Convolutional Neural Networks (CNN)

A feed-forward neural network is similar to a CNN. Data are processed to provide valuable results. CNN uses backpropagation, as well as semi-structured data [31], to train networks. Except for the input nodes, every node in the web is a processing element or a neuron. During “training,” the output and hidden layer weights are changed. Two successive layers cannot be fully connected. Unnecessary consequences may be removed after or during learning. Figure 2 shows a CNN network with hidden layers. Multilayer perceptrons offer two advantages over other varieties.

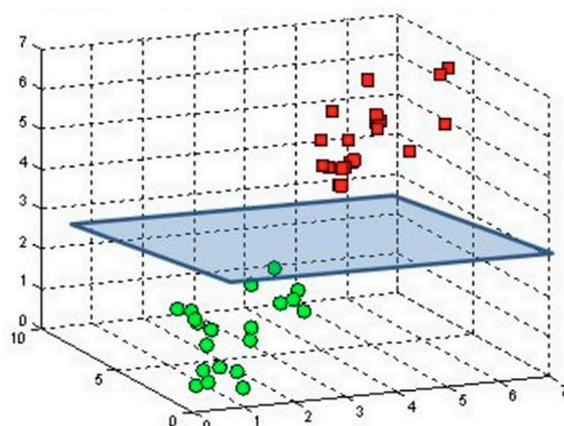


**Figure 2.** A CNN with N Inputs for precision agriculture.

Discriminant analysis approaches multilayer perceptrons to make no assumptions about how data are arranged or how covariance matrices of categorizable data sets are presented. Second, multilayer perceptrons allow for the creation of disconnected and nonsequential regions.

### 3.4. Support Vector Machines

SVMs are a “supervised machine learning” approach used for classification. This is machine learning. It develops an objective function using Vapnik-Chervonenkis (VC) [32] theory and structural risk reduction, and then it finds a partition hyperplane that meets class requirements. A SVM training approach organizes incoming data into one of two classes using a series of training datasets. Using two classifications, this model contains incoming data. SVM operates binary division to establish the border between two classes, such as Positive-class 1 and Negative-class 2. Figure 3 shows two classes, with the hyperplane in between. This minimizes the distance between the plane and each data point. The nearest data points are used to create the support vectors.



**Figure 3.** Classification of SVM with Hyperplanes as Class 0 and Class 1.

### 3.5. Two-Hierarchical Hybrid Ensemble Model

Figure 4 illustrates how the two-level soil moisture forecasting ensemble model presented is constructed from individual blocks and how it operates. And the working of the proposed algorithm is elaborated below.

The algorithm of the proposed approach is given below.

Input: Original data set  $D_o$  and the Preprocessed Dataset  $D_p$ .

Set of heterogeneous first level learning classifiers  $L_f = \{GPM, CNN, SVM\}$

Second-level learning algorithm  $L_s = \{ZeroR\}$

Start

Prepare the dataset for ensemble training and testing.

Training dataset—Dataset used for ensemble training (TR). This dataset contains soil moisture data.

Testing dataset—Dataset used for ensemble testing (TE). This dataset contains soil moisture data.

for  $i = 1$  to  $n$

//where  $n$  represents the number of heterogeneous classifiers

$h_{ti} = L_{fi} (TR)$

Train an individual first-level learner by applying the corresponding learning algorithm and the training data set

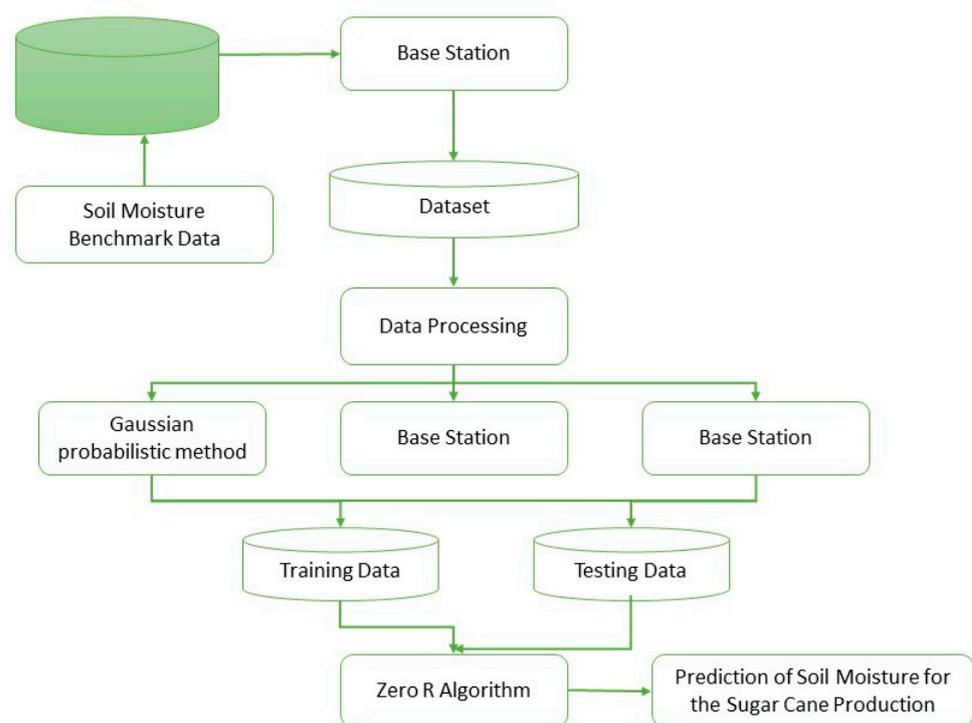
end

The output of base classifiers

Test the output from step 4 by using the second-level learning algorithm  $L_s$ .

Output the predicted soil moisture value

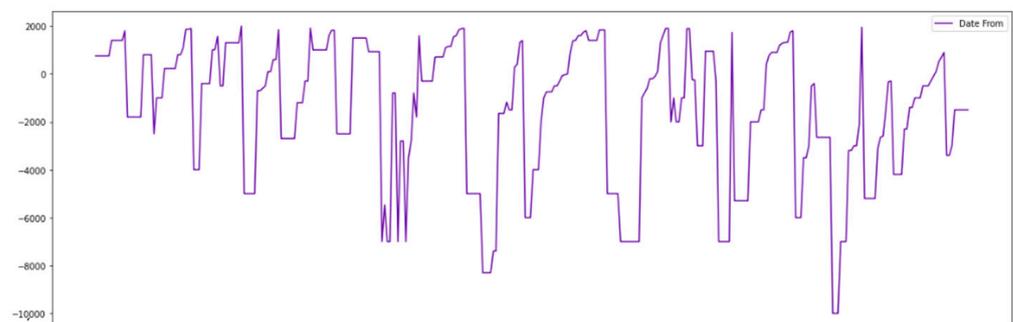
Stop



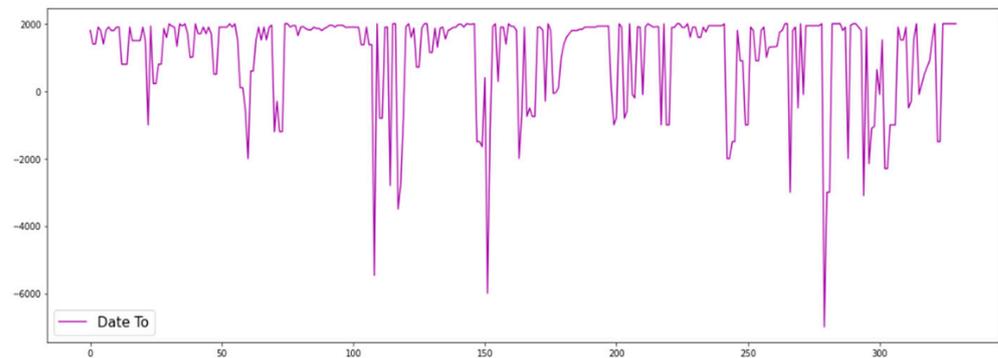
**Figure 4.** The proposed model for soil precision in agriculture using Hybrid Algorithm.

### 3.6. Godavari Plateau Delta Region Dataset

The research used soil moisture measurements from the Godavari River Plateau (<https://apwrims.ap.gov.in>) (accessed on 10 November 2022) and Krishna River Plateau (<https://krmb.gov.in/krmb/home>) [33], which involved online databases for both. The Krishna River plateau Soil Moisture Monitoring Network provides long-term soil moisture time data from 18 stations. The average of twice-monthly measurements is used. Figure 5 shows monthly data from 2016 to 2021 from a 10–30 cm layer at Station 1. Godavari river plateau Creek drains 600 km<sup>2</sup>. Fourteen stations detect soil moisture in the Godavari River plateau Creek’s basin. At 30-min intervals, soil moisture is monitored and averaged. Figure 6 shows the average daily soil moisture in the top 30 cm of station K4’s profile from 3 January 2022, to 31 August 2022.



**Figure 5.** Shows the soil moisture of the Godavari Plateau Region.



**Figure 6.** Krishna River plateau Soil Moisture.

### 3.7. Performance Evaluation Metrics of the Proposed Model

To determine how well the proposed model functions, statistical metrics, such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute % error (MAPE), which are shown in Equations (1)–(3), are utilized.

$$MAE = \frac{\sum_{i=1}^n |xi' - xi|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (xi' - xi)^2}{n}} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(xi' - xi)}{xi} \right| \times 100 \quad (3)$$

Where

$xi$  = actual\_value

$xi'$  = Predicted\_Value

$n$  = number\_of\_observations

The main objective of the function was to decrease the cost function between the actual and predicted values.

## 4. Results and Discussions

The Keras toolkit was used in the simulation-based testing that was carried out. This open-source program can predict time series and provides a machine-learning library that may be used for data mining jobs. As a consequence of this, this instrument is utilized to evaluate how well the proposed plan works. It is essential to partition the initial dataset into training and test sets before attempting any model evaluation. In this scenario, 70% of the dataset is put toward the training process, while the remaining 30% is used for testing.

Concerning the soil moisture datasets from the Krishna River plateau and the Godavari River plateau, tests were run with GPM, CNN, SVM, single-level GPM ensemble, single-level CNN ensemble, single-level SVM ensemble, and the proposed two-level soil moisture prediction model. Additionally, GPM, CNN, and SVM ensembles were tested individually.

The bagging approach generates single-level ensembles from GPM, CNN, and SVM classifiers. Examples of these ensembles are the GPM ensemble, the CNN ensemble, and the SVM ensemble. Figures 7 and 8 show the intended training and testing data projections for the Krishna River plateau data set when it was in the training phase and while it was in the testing phase, respectively. Figures 9 and 10 illustrate how a prediction was derived from the training and test data in the Godavari River plateau dataset.

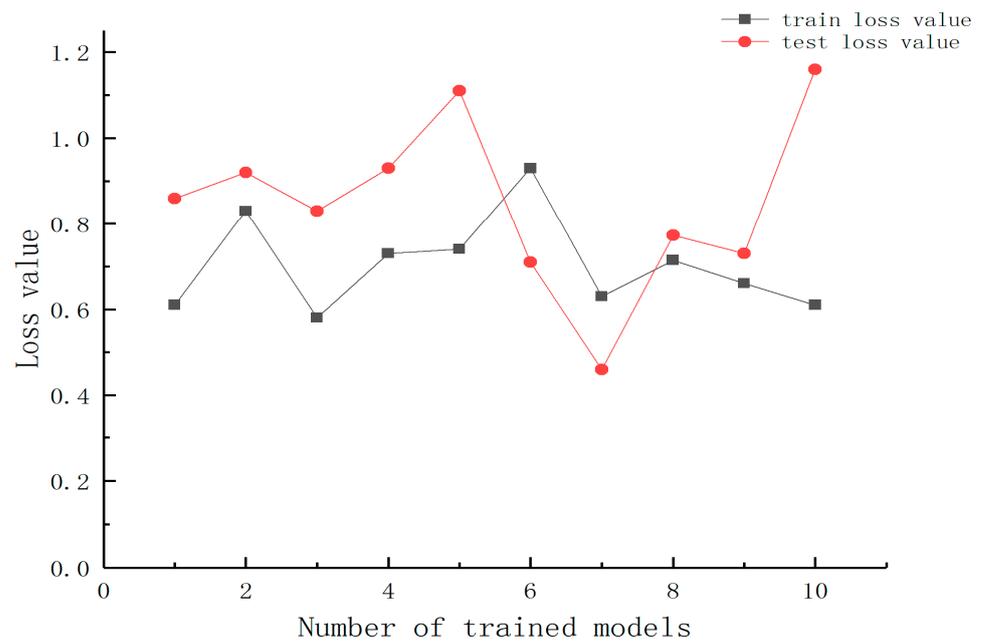


Figure 7. Forecast of training data—Godavari River plateau dataset.

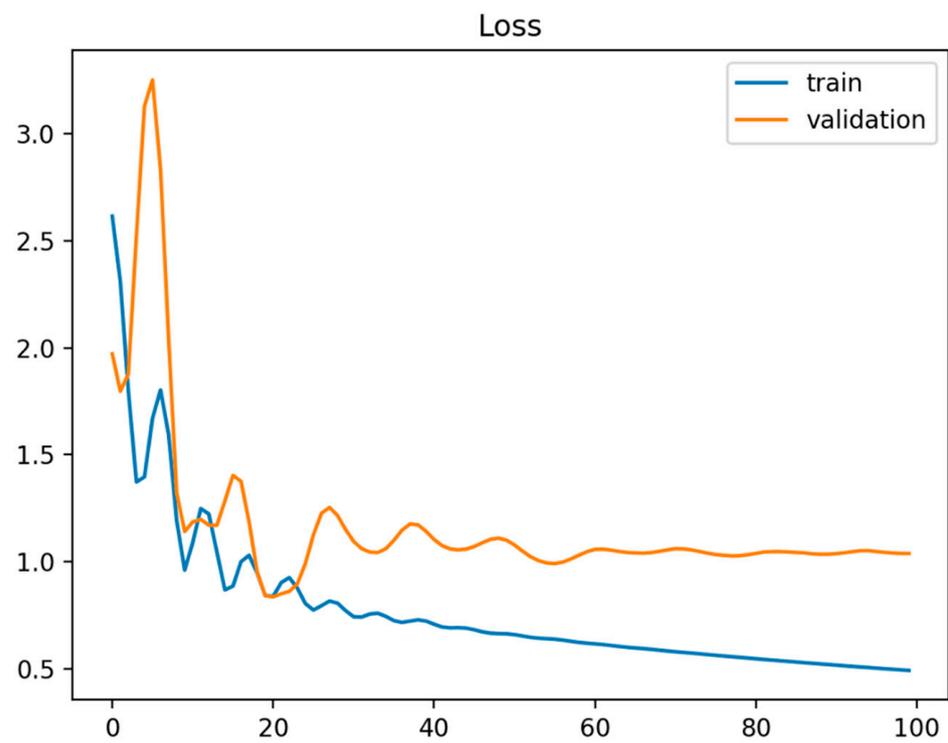


Figure 8. Forecast of testing data—Krishna River plateau dataset.

The one-step forward estimates of the future are shown in Figures 11 and 12 for two different datasets.

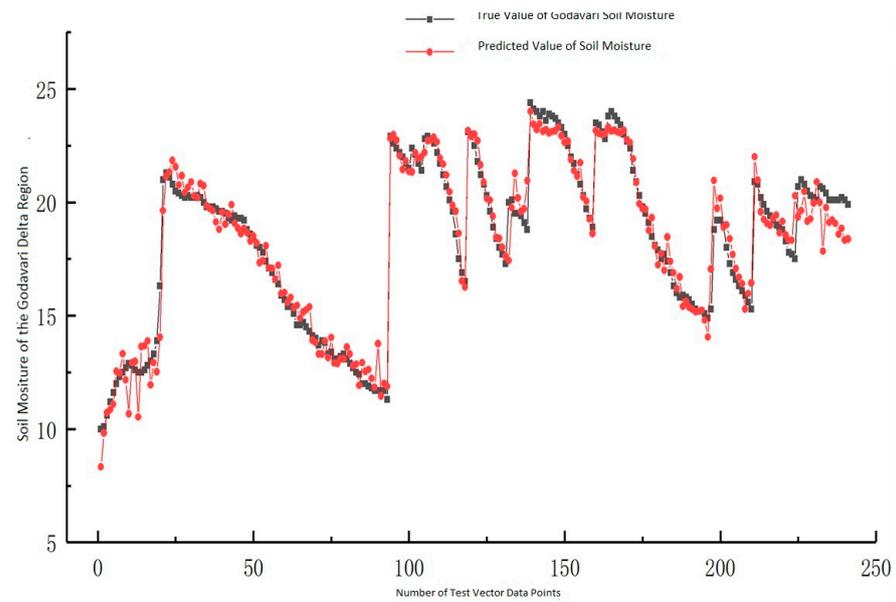


Figure 9. Forecast of training data—Godavari soil Moisture Dataset.

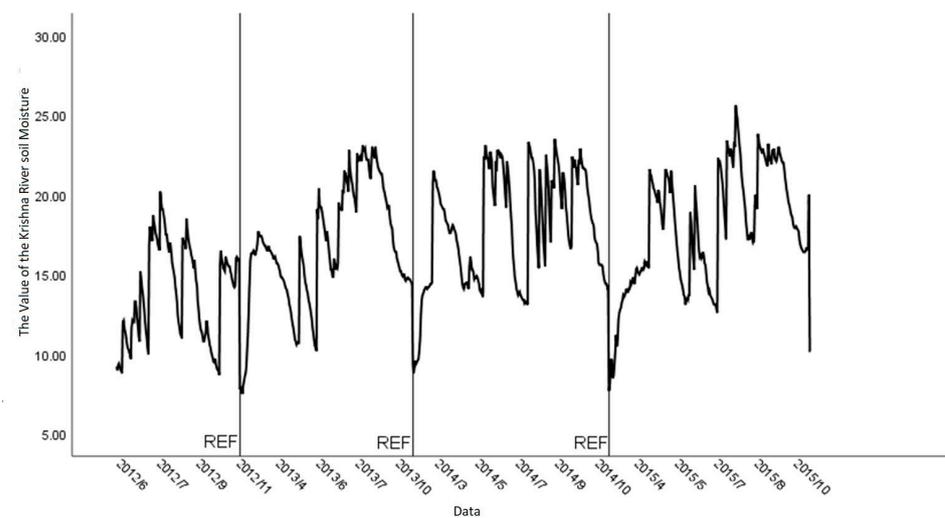


Figure 10. Forecast of testing data—Krishna Soil Moisture Dataset.

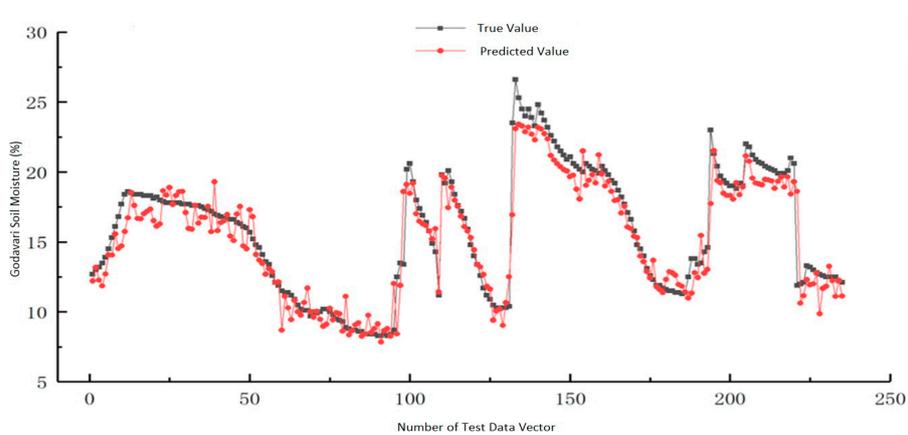
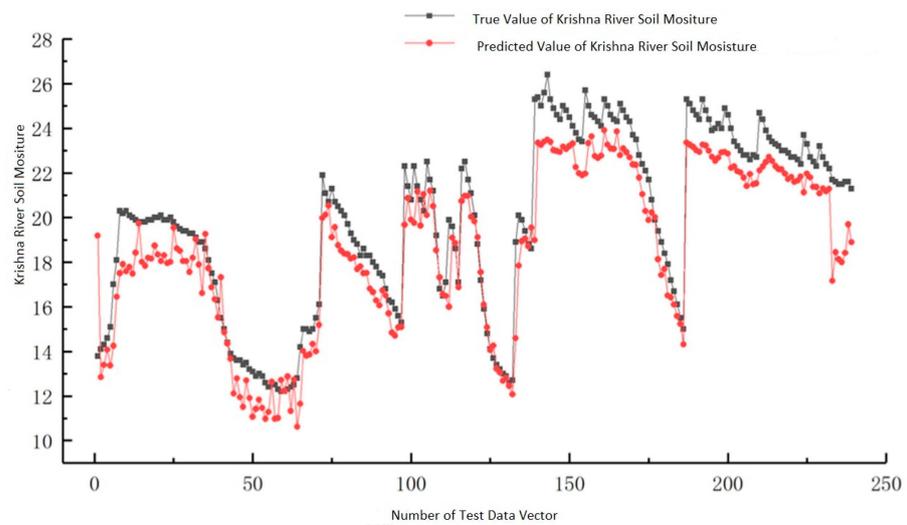


Figure 11. Future forecast of testing data—Godavari Soil Moisture Dataset.

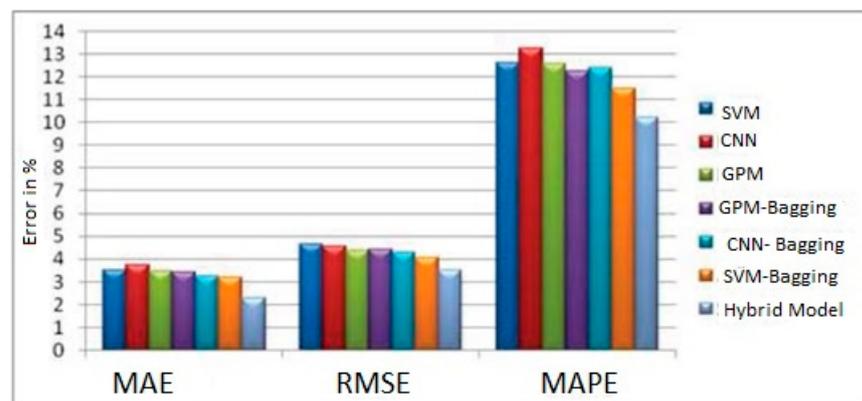


**Figure 12.** Future forecast of testing data–Krishna Soil Moisture Dataset.

Each machine learning method was applied to the same datasets to determine how accurately the suggested ensemble model could categorize items. Examples of individual classifiers are GPM, CNN, and SVM, in addition to the ensembles that go along with each of these. Using prediction accuracy, mean absolute error (MAE), root mean squared error (RMSE), and mean absolute % age error, the performance of the proposed model was compared to that of current machine learning approaches. Table 1 compares the effectiveness of several classifiers about the version of the one selected for use with the soil moisture data from the Krishna River plateau and Godavari River plateau. This is seen in Figures 13 and 14.

**Table 1.** Performance of the various classifiers with the proposed model.

| Method      | Krishna River Plateau Dataset |      |       | Godavari River Plateau Dataset |      |      |
|-------------|-------------------------------|------|-------|--------------------------------|------|------|
|             | MAE                           | RMSE | MAPE  | MAE                            | RMSE | MAPE |
| SVM         | 3.51                          | 4.62 | 12.65 | 0.72                           | 0.84 | 3.36 |
| CNN         | 3.77                          | 4.57 | 13.29 | 0.77                           | 0.86 | 3.82 |
| GPM         | 3.47                          | 4.35 | 13.6  | 0.76                           | 0.84 | 3.75 |
| GPM-Bagging | 3.43                          | 4.42 | 12.61 | 0.77                           | 0.98 | 4.06 |
| CNN-Bagging | 3.26                          | 4.27 | 12.42 | 0.78                           | 0.79 | 3.36 |
| SVM-Bagging | 3.24                          | 4.12 | 12.52 | 0.69                           | 0.81 | 3.29 |



**Figure 13.** Performance of the various classifiers with the proposed model.

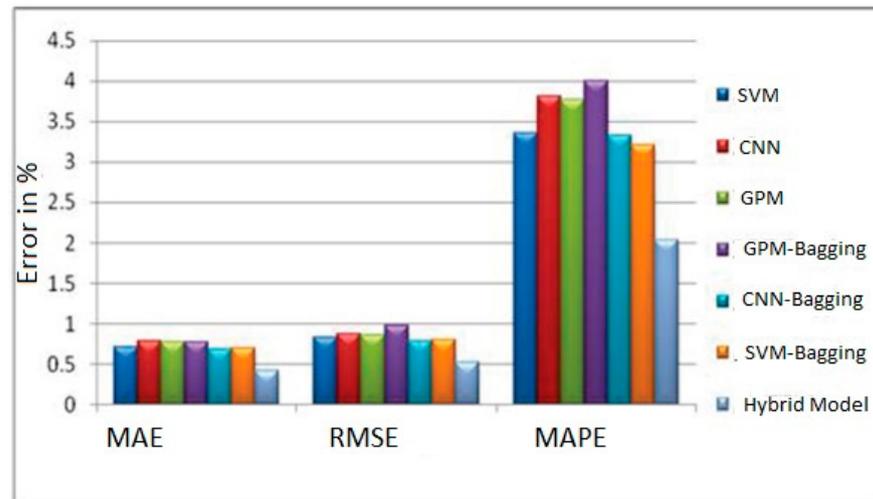


Figure 14. Performance of the various classifiers with the proposed model.

The ability of the two sets of data presented in Table 2 to accurately anticipate measurements is seen in Figures 15 and 16. The number of various classifiers in the testing set that could predict the one displayed accurately is one way to determine how accurate a measurement’s prediction is.

Table 2. Prediction accuracy of various classifiers.

| Classifier/Result | Prediction Accuracy        |                             |
|-------------------|----------------------------|-----------------------------|
|                   | Krishna River Plateau Data | Godavari River Plateau Data |
| SVM               | 76.89%                     | 81.23%                      |
| CNN               | 74.22%                     | 71.98%                      |
| GPM               | 75.36%                     | 77.78%                      |
| GPM-Ensemble      | 77.96%                     | 78.92%                      |
| CNN-Ensemble      | 76.96%                     | 75.69%                      |
| SVM-Ensemble      | 79.86%                     | 83.59%                      |
| Hybrid Model      | 86.97%                     | 92.69%                      |

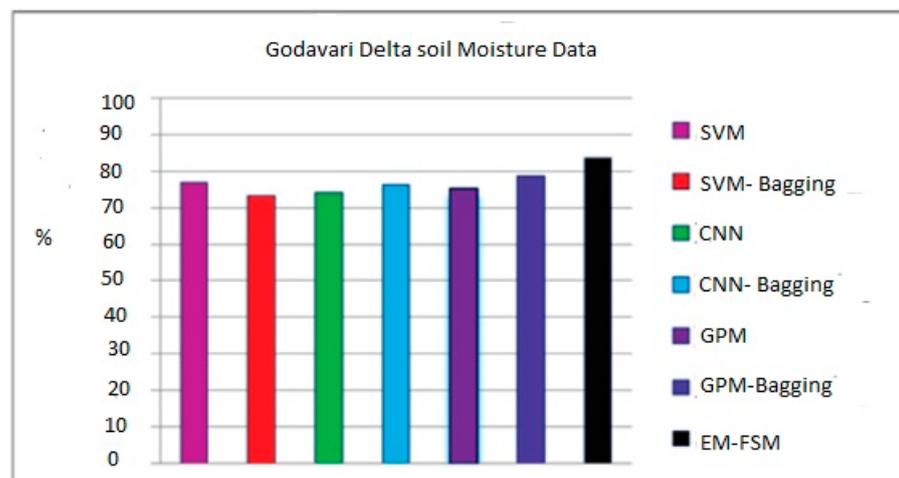
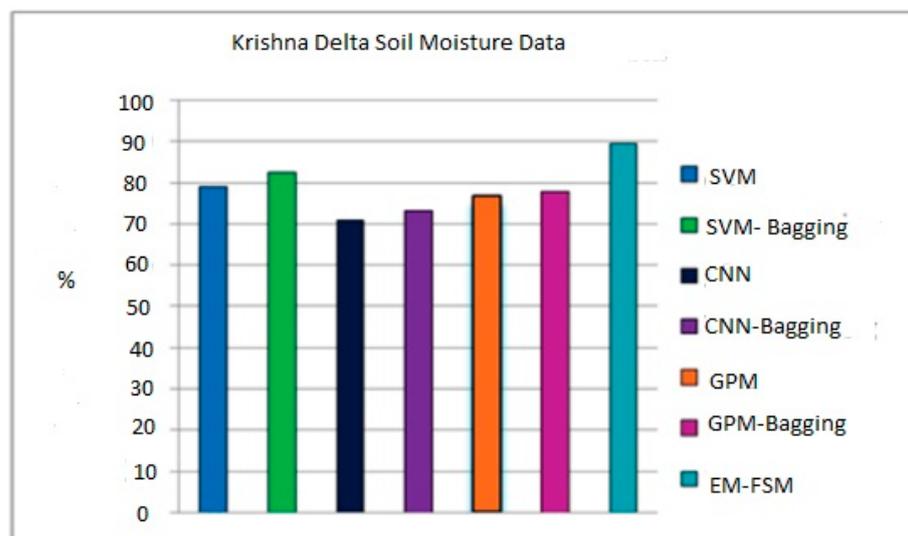


Figure 15. Prediction accuracy of different classifiers.



**Figure 16.** Prediction accuracy of different classifiers for the Godavari River plateau.

Compared to other individual classifiers, the SVM classifier has an accuracy ranging between 76.85% and 79%. The accuracy of the SVM-bagging ensemble is 78.81% and 82.56% higher than that of other bagging ensembles. Still, the accuracy of the proposed two-level ensemble model is 83.52% and 89.56% higher than that of all other ensemble models. The previously used solo classifiers and the single-level ensemble classifiers constructed using GPM, CNN, and SVM outperform the newly presented ensemble model. According to this research's findings, ensemble neural networks to forecast soil moisture data may result in more accurate forecasts. Using this model will assist farmers and others responsible for making policies in creating future irrigation plans. In this particular piece of research, a method for the long-term collection of data on soil moisture is investigated. The GPM, CNN, and SVM classifiers each provide their own set of helpful rules to the ensemble model that is being proposed. The experiment's findings demonstrate that, when several models are effectively connected, they are superior to individual learners in predicting soil moisture. Farmers can profit from this strategy, since it provides them with reliable information.

Alterations to the methods used to cultivate crops, as well as new knowledge on the moisture content of the ground in the years to come. The findings of the studies indicate that the suggested ensemble model accomplishes its objective of increasing prediction accuracy. The most fantastic accuracy of prediction was found to be 83.53% for both of the datasets utilized and 89.53%, respectively. In addition to these applications, soil moisture estimates may be used to plan irrigation, calculate agricultural productivity, provide early warning of drought conditions, and forecast runoff.

## 5. Conclusions

It has been observed that sugarcane growth can be forecasted by looking at the pedogenetic traits of the soil moisture. The GPM helps to learn more about the different qualities of soil and the quickest ways to handle sugarcane. Using machine learning in sugarcane output was interesting because it allowed farmers to take advantage of traits that could be taught rapidly and cheaply.

Regarding the classification process, the hybrid approach recommended for the hybrid model for learning has an accuracy rate of 89.53%. The proposed system's efficiency was evaluated and contrasted with the efficiency of various other machine learning methods. The findings demonstrate that the strategy that was recommended is effective.

## 6. Future Study

The following alterations are currently in the works. Installing real-time sensor nodes across agricultural areas to monitor various factors for research purposes is important. More study is required to determine which route design will be most effective for farming data processing applications. Using the methodology outlined in this research work, researchers have conducted studies on crops that are more economically significant. To investigate and assess the performance of various hybrid evolutionary computing technologies to find a solution to the existing challenge is important.

**Author Contributions:** Conceptualization, E.S.N.J.; methodology, D.B.; software, E.S.N.J. and D.B.; validation, T.-h.K. and E.S.N.J.; formal analysis, N.T.R.; writing—original draft preparation, E.S.N.J., T.-h.K., D.B. and N.T.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data are publicly available and Benchmarked.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bhargavi, G.; Arunnehr, J. An extreme machine learning model for evaluating landslide hazard zonation in nilgiris district, causative factors and risk assessment using earth observation techniques. In *Lecture Notes on Data Engineering and Communications Technologies*; Springer Nature Singapore: Singapore, 2023. Available online: [www.scopus.com](http://www.scopus.com) (accessed on 23 November 2022). [CrossRef]
2. United Nations World Meter. Available online: <https://www.worldometers.info> (accessed on 10 April 2023).
3. Huang, Y.; Lan, Y.; Thomson, S.J.; Fang, A.; Hoffmann, W.C.; Lacey, R.E. Development of soft computing and applications in agricultural and biological engineering. *Comput. Electron. Agric.* **2010**, *71*, 107–127. [CrossRef]
4. Asha, R.; Babu, G.S.K.; Teja, T.S. Production and marketing of sugarcane in Visakhapatnam district of Andhra Pradesh. *J. Res. Angraui* **2019**, *47*, 69–77. Available online: <https://epubs.icar.org.in/index.php/TJRA/article/view/133728> (accessed on 23 November 2022).
5. Singh, P.; Pandey, P.C.; Petropoulos, G.P.; Pavlides, A.; Srivastava, P.K.; Koutsias, N.; Deng, K.A.K.; Bao, Y. Hyperspectral remote sensing in precision agriculture: Present status, challenges, and future trends. In *Earth Observation, Hyperspectral Remote Sensing*; Pandey, P.C., Srivastava, P.K., Balzter, H., Bhattacharya, B., Petropoulos, G.P., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 121–146. ISBN 9780081028940. [CrossRef]
6. Singh, D.; Jain, N.; Jain, P.; Kayal, P.; Kumawat, S.; Batra, N. PlantDoc: A Dataset for Visual Plant Disease Detection. *arXiv* **2019**, arXiv:1911.10317. [CrossRef]
7. Anas, M.; Liao, F.; Verma, K.K.; Sarwar, M.A.; Mahmood, A.; Chen, Z.-L.; Li, Q.; Zeng, X.-P.; Liu, Y.; Li, Y.-R. Fate of nitrogen in agriculture and environment: Agronomic, eco-physiological and molecular approaches to improve nitrogen use efficiency. *Biol. Res.* **2020**, *53*, 47. [CrossRef] [PubMed]
8. Kai, P.M.; de Oliveira, B.M.; de Costa, R.M. Deep Learning-Based Method for Classification of Sugarcane Varieties. *Agronomy* **2022**, *12*, 2722. [CrossRef]
9. Nihar, A.; Patel, N.R.; Danodia, A. Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India. *J. Indian Soc. Remote Sens.* **2022**, *50*, 1519–1530. [CrossRef]
10. Militante, S.V.; Gerardo, B.D. Detecting Sugarcane Diseases through Adaptive Deep Learning Models of Convolutional Neural Network. In Proceedings of the 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 20–21 December 2019; pp. 1–5. [CrossRef]
11. Wang, J.; Guo, J.Y. Sugarcane cultivation management technology. *Agric. Technol. Serv.* **2013**, *30*, 321–322.
12. Everingham, Y.; Sexton, J.; Skocaj, D.; Inman-Bamber, G. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* **2016**, *36*, 27. [CrossRef]
13. Talaviya, T.; Shah, D.; Patel, N.; Yagnik, H.; Shah, M. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artif. Intell. Agric.* **2020**, *4*, 58–73. [CrossRef]
14. Blackmore, I.; Rivera, C.; Waters, W.F.; Iannotti, L.; Lesorogol, C. The impact of seasonality and climate variability on livelihood security in the Ecuadorian Andes. *Clim. Risk Manag.* **2021**, *32*, 100279. [CrossRef]
15. Kok, Z.H.; Shariff, A.R.M.; Alfatni, M.S.M.; Khairunniza-Bejo, S. Support Vector Machine in Precision Agriculture: A review. *Comput. Electron. Agric.* **2021**, *191*, 106546. [CrossRef]
16. Tavva Srinivas, M. Nedunchezhiyan The nexus between adoption and diffusion of production technologies with yield: Evidence from sweet potato farmers in India. *Technol. Soc.* **2020**, *60*, 101208. [CrossRef]
17. Sequeros, E.; Schreinemachers, P.; Depenbusch, L.; Shwe, T.; Nair, R.M. Impact and returns on investment of mungbean research and development in Myanmar. *Agric. Food Secur.* **2020**, *9*, 5. [CrossRef]

18. Uyeh, D.D.; Iyiola, O.; Mallipeddi, R.; Asem-Hiablie, S.; Amaizu, M.; Ha, Y.; Park, T. Grid Search for Lowest Root Mean Squared Error in Predicting Optimal Sensor Location in Protected Cultivation Systems. *Front. Plant Sci.* **2022**, *13*, 920284. [[CrossRef](#)] [[PubMed](#)]
19. Solomon, S.; Swapna, M. Indian Sugar Industry: Towards Self-reliance for Sustainability. *Sugar Tech* **2022**, *24*, 630–650. [[CrossRef](#)] [[PubMed](#)]
20. Tian, H.; Wang, P.; Tansey, K.; Zhang, J.; Zhang, S.; Li, H. An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agric. For. Meteorol.* **2021**, *310*, 108629. [[CrossRef](#)]
21. Kai, P.M.; Oliveira, B.M.; Vieira, G.S.; Soares, F.; Costa, R.M. Effects of resampling image methods in sugarcane classification and the potential use of vegetation indices related to chlorophyll. In Proceedings of the 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 12–16 July 2021; pp. 1526–1531.
22. Panwar, E.; Singh, D.; Sharma, A.K. Exploring the Possibility of Assessing Biochemical Variables in Sugarcane Crop with Sentinel-2 Data. In Proceedings of the 2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020), Virtual, 26 September–2 October 2020; pp. 3712–3715.
23. Shi, L.; Duan, Q.; Ma, X.; Weng, M. The Research of Support Vector Machine in Agricultural Data Classification. In Proceedings of the Computer and Computing Technologies in Agriculture V. CCTA 2011. IFIP Advances in Information and Communication Technology, Beijing, China, 29–31 October 2011; . Li, D., Chen, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 370. [[CrossRef](#)]
24. Abdel-Sattar, M.; Aboukarima, A.M.; Alnahdi, B.M. Application of artificial neural network and support vector regression in predicting mass of ber fruits (*Ziziphus mauritiana* Lamk.) based on fruit axial dimensions. *PLoS ONE* **2021**, *16*, e0245228. [[CrossRef](#)] [[PubMed](#)]
25. Lv, M.; Wei, H.; Fu, X.; Wang, W.; Zhou, D. A Loosely Coupled Extended Kalman Filter Algorithm for Agricultural Scene-Based Multi-Sensor Fusion. *Front. Plant Sci.* **2022**, *13*, 849260. [[CrossRef](#)]
26. Gałęzewski, L.; Jaskulska, I.; Jaskulski, D.; Lewandowski, A.; Szyplowska, A.; Wilczek, A.; Szczepańczyk, M. Analysis of the need for soil moisture, salinity and temperature sensing in agriculture: A case study in Poland. *Sci. Rep.* **2021**, *11*, 16660. [[CrossRef](#)]
27. Lu, J.; Tan, L.; Jiang, H. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture* **2021**, *11*, 707. [[CrossRef](#)]
28. Matko, V.; Koprivnikar, J. Quartz sensor for water absorption measurement in glass-fiber resins. *IEEE Trans. Instrum. Meas.* **1998**, *47*, 1159–1162. [[CrossRef](#)]
29. Eze, P.N.; Kumahor, S.K. Gaussian process simulation of soil Zn micronutrient spatial heterogeneity and uncertainty—A performance appraisal of three semivariogram models. *Sci. Afr.* **2019**, *5*, e00110. [[CrossRef](#)]
30. Cravero, A.; Pardo, S.; Galeas, P.; López Fenner, J.; Caniupán, M. Data Type and Data Sources for Agricultural Big Data and Machine Learning. *Sustainability* **2022**, *14*, 16131. [[CrossRef](#)]
31. Gey, S. Vapnik–Chervonenkis dimension of axis-parallel cuts. *Commun. Stat. Theory Methods* **2018**, *47*, 2291–2296. [[CrossRef](#)]
32. Andhra Pradesh Water Resources Management. Available online: <https://apwrims.ap.gov.in/> (accessed on 10 November 2022).
33. Krishna River Board Management. Available online: <https://krmb.gov.in/krmb> (accessed on 10 November 2022).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.