Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Predicting Crops based on Soil Features and its Surroundings Using the Capsule A-BiLSTM Method

Baseera [1], N.Suresh Kumar[2], Uppara Raghu Babu [3], Neelam Kumar [4], Swetha Reddy A [5], Venkata Ramana K[6]

[1] School of Computing Science and Engineering (SCSE), VIT Bhopal University, Bhopal-Indore Highway -18, Sehore, Madhya Pradesh, India,

[2] Department of IT, Sri Ramakrishna Engineering College, Coimbatore, India, nsuresh2@srec.ac.in

[3] Civil Engineering Department, Srinivasa Ramanujan Institute of Technology, Anantapur, India,

[4] Computer Department, Shree Ramchandra College of Engineering, Pune, Maharashtra, India,

[5] Department of Electronics and Communication Engineering, R.M.K. College of Engineering and Technology, Chennai, India,

[6] Department of CSE, QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India,

Abstract. Traditionally, farmers grew crops based on their knowledge and experience. The weather has changed, though, and it's hurting food yields. Because of this, farmers can't pick the right crop(s) based on the soil and the climate, and trying to guess which crop(s) would grow best on their land by hand has usually failed. More crops are grown when crop predictions are accurate. Here, machine learning is very important for making predictions about crops. Predicting crops relies on the soil, the location, and the weather. Choosing the right attributes for the right crop(s) is an important part of how feature selection methods make predictions. Putting together all of the features from the raw data without first checking what part they play in the model-making process will make it more complicated than it needs to be. Also, adding traits that don't add much to the DL model will make it more time and space-consuming and change how accurate the results are. The suggested method includes preparation, choosing features, and teaching the model. The suggested method uses SG Denoising for preprocessing. CFS and MIFS are used in feature selection. The last step is to use Cap-A-BiLSTM to train the model. When compared to two other common methods, the proposed solution works very well. The results show that a Cap-A-BiLSTM can make more accurate predictions than the current classification technique.

Keywords: Savitzky – Golay Denoising (SGD) · Correlation Feature Selection (CFS) · Mutual Information-based Feature Selection (MIFS).

1 Introduction

Predictions of crop growth and soil processes can provide timely information for management recommendations. Such information can improve marketing choices, environmental quality, and profitability. Efforts are underway to predict agricultural yields by season by utilizing field surveys, expert opinion, remote sensing, statistical models, and process-based simulation models. The degree to which you require accuracy, transparency, or fine-grained control over your results may dictate which approach is best for your situation. Although farmer surveys can provide valuable information, they can only report on past events and rely on the participation of willing volunteers. The critical underlying processes for environmental performance cannot be adequately described by statistical models and remote sensing due to their descriptive nature. Conversely, process-based models can shed light on the mechanisms at work in soil and crops, which in turn affect both yield and environmental impacts. A potential drawback of these models is the large amount of data they need for input has shown that using a combination of explanatory and predictive capabilities to evaluate adaptive management choices can be promising. Predicting yield and soil-crop dynamics during the growing season is challenging because of the intricacies of weather-related uncertainty, soil-crop genetic variability, and management strategy variability. Crop production is affected by a lot of different factors, such as the genotype of the crop, the environment, and the methods used for management. Companies that sell seeds have come a long way in terms of improving crop genotype. Because environmental conditions can fluctuate in both geography and time, crop yield varies substantially from year to year and from one location to another. So, accurate yield forecasting is fantastic for food production all around the world. Accurate predictions enable quick decisions about imports and exports. Various environmental factors, including the weather accurate estimation is sometimes hindered by components that display complex nonlinear effects. The employment of random forests, regression trees, association rule mining, artificial neural networks, and multivariate regression has been the focus of many machine learning research aimed at predicting agricultural output. Agriculture accounts for a sizable chunk of India's gross domestic product. The idea that India is an abundance paradise is a common misconception. Farming employs half of India's labor force. When it comes to farming, soil is king. Nevertheless, conventional farming methods are still used by farmers in the modern day. Since farmers did not get satisfactory results using the conventional method, the quantity of crops is not increasing. Having good soil is essential if you wish to increase your crop yield. The soil is tested. Therefore, soil testing is vital and should not be left to farmers alone. A lot of study has focused on crop models as a tool for making predictions. As an example, compared two crop models' ability to predict yields within a season. One model used mean climate data, whereas the other used stochastically produced climatic data. With and without the incorporation of weather forecast data, used APSIM to forecast phenology and yields for soybeans and maize, respectively. There are a lot of instances in the literature of people using crop modelling to anticipate various components of the cropping system. It is also unclear whether the numerous studies that have using machine learning to predict agricultural output in specific areas can be applied to other areas. For instance, used real data collected for one aim, which may not be generalizable to different crops or locations. However, some have restricted their reusability owing to design decisions that are specific to crops and regions, while others have utilized publically available climate and satellite data. This study aims to address the challenge of understanding the value of various data sources, characteristics, predictors, and machine learning techniques for various crops across diverse locations and periods. We will achieve this by implementing modular and reusable workflows. Reusable workflows allow scientists to consistently use the same set of input data for multiple tests across different crops and countries, allowing them to make predictions about the beginning or end of the season. The models might be adjusted for specific crops and areas by incorporating additional data sources, higher-level characteristics, and other optimizations. The MARS Crop Yield Forecasting System (MCYFS), operated by the Joint Research Centre (JRC) of the European Commission and the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA), is one example of a large-scale crop yield forecasting system that can be used to build and assess models for predicting crop yields in different regions and with different crops. However, operational systems that are familiar to people do not involve machine learning. The statistical models are built using data gathered from many sources such as weather observations, field surveys, crop growth models, remote sensing indicators, and yield statistics.

2. Literature Survey

Soil status monitoring is a hot commodity in precision agriculture because it allows farmers to fine-tune their tillage, fertilization, and irrigation strategies. A thorough understanding of the soil's characteristics is crucial for effectively applying soil management operations, procedures, and treatments and for assisting producers in making educated farming decisions [1]. Wet chemistry and other conventional analytical methods can be prohibitively expensive, time-consuming, and equipment-intensive when dealing with dense geographical samples [2]. A faster, nondestructive, more eco-friendly, repeatable, and reproducible alternative to the traditional wet chemistry method is soil Visible and Near-InfraRed (V-NIR) reflectance spectroscopy [3]. Various studies have shown that V-NIR spectra are reliable for evaluating the physical, chemical, and biological aspects of soil, although their relative Data is rarely collected with the intention of making comparisons, contrasts, and critical evaluations. Several assessments have found that reflectance spectroscopy can be a useful tool for forecasting soil physical parameters none of them followed the strict protocol of a "systematic review," as far as the writers are aware, when it came to the chemical and biological properties [4]. Importantly, a "systematic review" removes the subjectivity that comes with choosing which works to include or exclude, which is a flaw in conventional literature reviews that could skew the findings. Several restrictions are included in the current literature. Soil organic matter [5] and soil carbon content are two examples of properties that are typically reviewed in isolation. Furthermore, the majority of reviews compare V-NIR and Mid-InfraRed (MIR) spectroscopy data. "Machine Learning Algorithms for Efficient Crop Yield Prediction" Agricultural yields were categorized according to productivity using artificial neural networks in this study, which allowed for batch processing[6]. It will also specify the range of production. Using regression analysis, we may calculate the actual crop production and the projected cost. "Weather forecasting using machine learning" [7] is considered in order to compile historical weather data from various weather stations for the goal of making weather predictions. "Rainfall prediction using Machine Learning Techniques" [8] is a good read for anyone interested in weather forecasting for crop prediction. To understand how to predict crop yields using weather and soil variables, see "Prediction of Crop Yield using Machine Learning" [9]. "Sugarcane Crop prediction Using Supervised Machine Learning" aims to forecast the specific crop by combining and applying descriptive analytics to three datasets: the Soil dataset, the Rainfall dataset, and the Yield dataset. [10] proposes a semi-parametric deep neural network model in his 2006 study "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture" for the purpose of predicting crop yields and evaluating the effects of climate change. By assessing soil, environmental, and abiotic variables, employed multiple ML algorithms to predict groundnut yields in their study "Groundnut Prediction Using Machine Learning Techniques". Handheld and vehicle-mounted proximate sensing techniques show significant promise for monitoring soil and crop variables that could be responsible for changes in agricultural output. Statistical and neural methods for site-specific yield prediction[11] Proper sensing allows for the collection of ground-truth data, which may be used by Precision Agriculture to map cultivated areas and enable site-specific management of irrigation, herbicides, and fertilizers in real-time. [12] When combined with state-of-the-art data processing methods like ML, these precision agriculture technology may extract crucial information that determines crop productivity. Because potatoes are a very variable crop and because of the many management techniques employed to keep it under control, site-specific management is an option. [13] However, at this time, we do not have a good understanding of how easily measurable physiochemical parameters affect yield in potato fields. Many methods were suggested by [14] to determine the impact of soil quality on harvest success. One method is to use mechanistic growth models, which aren't perfect. Among these methods is the investigation of large datasets, such as those employed in precision agriculture? Various data analysis techniques have been used by numerous researchers to investigate this idea. [15] The third strategy makes use of agronomic practices, including as data collection over multiple site-years. This method might be the most time-consuming, but it might also yield the greatest results. In order to choose characteristics from a dataset, [16] suggested a hybrid combination approach that uses the Particle Swarm Optimization - Support Vector Machine (PSO-SVM). Multiple benchmark datasets were subjected to this method's testing. [17]states that conducting functional phylogenetic analysis for large-scale genetic screening is a challenging endeavor; nonetheless, it can be utilized to capitalize on inherent variability in photosynthesizing ability and enhance yields. Researchers looked at how well leaf reflectance spectroscopy could predict photosynthetic efficiency specs in two different types of seeds—C4 Zea mays and C3 Brassica oleracea—and found that phenotyping leaf reflectance improved the photosynthetic ability of both types of crops. [18] Proposed a Boruta-based wrapper feature selection method for crop prediction. The method improves the standard of the predictors as well as their ability to make accurate predictions. The Z score gives the most accurate assessment in Boruta since it takes into consideration the spread of the mean accuracy loss across forest trees. [19]Due to the crops it generates, this industry ranks among the most significant for the nation's economy. Once again, agricultural output is highly dependent on a country's climate, soil type, and other climatic factors. Accordingly, evaluating the relationship between these indicators is vital for optimizing production and forecasting crop yield. [20] The available land type is an important factor to think about because different types of terrain are ideal for cultivating specific crops. Because of the favorable soil conditions for the cultivation of wheat, jute, T-Aman, and mustard, our research primarily focuses on medium highland and highland terrain types [21]. Then, for each of these land types, we ran two well-known data mining analyses—clustering and regression—on their own sets of variables. Soil health is maintained by a complex network of subsurface processes in both natural and agricultural settings. The ability of soil to maintain clean water and air around it and to increase plant and animal productivity is called soil quality [22]. Therefore, healthy soils are essential for the agricultural and pastoral industries, which play a significant role in ensuring a steady supply of food and a healthy economy. Soil bacteria perform a variety of functions, including nutrient cycling, impacting plant growth, and potentially causing or preventing illnesses [23]. Macro organisms collaborate with bacteria for this purpose, but they also independently contribute significantly to processes such as decomposition. Nutrient levels, metal contaminants, and soil structure are examples of abiotic factors that are typically monitored by soil quality programmes, despite the fact that living creatures are vital to maintaining healthy soil ecosystems [24]. Biological measures, such as soil respiration or microbial biomass, are sometimes used in monitoring efforts; however, these measurements tend to be broad and crude. However, there are many who find that species-specific indicators, such as earthworms, are more reliable [25]. In addition to providing important data on the biological activity of the ecosystem, soil organisms react only to bioavailable nutrients and contaminants, rather than chemical assays that capture the total amount. Incorporating biological indicators into soil monitoring allows for a more sensitive, relevant, and thorough understanding of the consequences of human activities on soil, since soil bacterial communities are extremely sensitive to changes in soil conditions. The diversity and composition of bacterial communities fluctuate in response to variations in soil acidity. Many consider this to be the most crucial explanatory variable when studying bacterial community richness on a national or even international level. Actually, pH measurements alone can predict bacterial diversity on a wide scale. Soil type, nutrient concentrations, plant diversity, and soil moisture are all variables that can affect bacterial community composition, as has been shown in previous studies.

3. Proposed System

Agriculture is an area of study that is growing. In agriculture, crop prediction is especially important and depends a lot on things like soil and environment factors like temperature, humidity, and rainfall. Farmers used to be able to choose what crops to grow, keep an eye on their progress, and know when the crops were ready to be picked. Fast changes in the environment, on the other hand, have made it hard for farmers to keep doing what they're doing. So, in recent years, machine learning methods have taken over the job of making predictions. Several of these techniques were used in this work to figure out crop yield.

3.1 Preprocessing

Nonstructural components that can cause strain changes are removed from the data by pretreatment before it is used in the damage detection analysis process. The data is filtered and/or zeroed as part of the preparation step. Each component has been fine-tuned to cancel out any impact from dynamic effects, temperature effects, or fluctuations in actual load.

3.1.1 Savitzky – Golay Denoising

For pre-treatment of dirt spectra, the SG smoothing filter is often used. It's a low-pass filter that gets rid of all the high-frequency noise in the spectrum while letting the low-frequency data pass. It also uses the weighted-average method for the moving window and is based on the least squares fitting of a curve local polynomial. Its weighting coefficient, on the other hand, is not a simple constant window. Instead, it is found by fitting the least squares of a given higher-order equation into a sliding window [22]. The main idea behind it is to get the rebuilt curve closer and closer to the upper envelope of the original curve over time. Using the SG method for smooth filtering-based denoising could make the spectrum smoother and lower the amount of noise interference.

This is the phrase for SG filtering:

$$A_d^* = \frac{\sum_{b=-q}^{q} X_b A_d}{P} \tag{1}$$

In this case, A_d^* represents both the original and rebuilt spectral data, whereas P is the number of datapoints in the sliding window, which is equal to $P = 2q + 1$. The pane of glass measures 2 metres plus one. The filter window width and the order of the polynomial utilised for smooth fitting are the two real-world factors required for SG filtering. Adjusting the width of the filter window alters the smoothing effect. A broader window results in a more uniform spectrum. Similarly, the order of the fitting polynomial impacts the screening results. As the order rises, the fit becomes better. We used a fitting polynomial with an order of 2 and a filter window size of 21.

3.2 Feature Selection

3.2.1 CFS

One of the best ways to get the most linked features is to use correlation-based feature selection. A lot of data scientists have used this method to get features for their machine learning systems. Our set of data on corn and soybean beans shows that the expected yield is linked to things like weather, soil, and management. Many common types of association are used to choose features. In order to get mostly correlated features that were linked to yield, we used Pearson correlation in our regression model. To find Pearson's correlation coefficient, divide the covariance of two variables by the product of their standard deviations. This is how the coefficient is shown mathematically:

$$\sigma_{Y,I} = \frac{cov(Y,I)}{\rho_Y \rho_I} \tag{2}$$

that is, $\sigma_{Y,I}$ is the Person correlation coefficient between Y and I. For two traits Y and I, $cov(Y,I)$ shows how much they are alike. The difference between traits Y and I is shown by $\rho_Y \rho_I$.

To use correlation to choose features, the crop yield data set was split into training and testing parts. The data was then pre-processed for feature engineering by filling in missing values, and then highly correlated features were calculated. This leads to dimensionality reduction or the extraction of the most correlated features from the original dataset. The features are then fed into the right DL algorithms for crop prediction.

3.2.2 MIFS

One powerful way to find feature relationships between datasets is to use MIFS. It is also known as the removal process because it reduces the size of the dataset's input by finding the features that are related to each other without changing the most important features that help with the classification or regression problem. You can choose features by figuring out how the random features rely on each other [23]. This dependence should always be symmetric and not negative. It should only pick the independent traits that show a value of zero.

There is mutual information between two sets of discrete random traits $Y = y_1, y_2, \dots., y_v$ and $I = i_1, i_2, \dots., i_v$. This is shown by

$$B(Y,I) = \sum_y \sum_i n(y,i) \log \frac{n(y,i)}{n(y)n(i)} \tag{3}$$

where $B(Y,I)$ is the information that two discrete random traits Y and I share with each other. The values of the discrete traits Y and I are shown by $(y_1, y_2, \dots., y_v)$ and $(i_1, i_2, \dots., i_v)$. The joint density function is shown by $n(y,i)$. The functions $n(y)$ and $n(i)$ show the marginal density. When using mutual information to choose features, the training and testing data from the first dataset were first cleaned up and filled in with missing values as part of feature engineering. Then, the mutual information feature was calculated. So, the original dataset's dimensions are cut down, and then the features that were extracted are fed into the right DL algorithms for predicting crop yield.

3.3 Performance evaluation of the model

This piece talked about BiLSTM-Capsule, a capsule network framework that works with BiLSTM. The framework is made up of two parts: a BiLSTM module and an attention layer and a capsule combination module.

3.3.1 BiLSTM-C

The input is computed by the first BiLSTM layer, which then sends the same image to the capsules. By making hidden vectors as output, the BiLSTM stores any text instance that is shown as a dense vector. These secret vectors are what all the capsules sense. The boxes that are most likely to be in that state will be "enabled," and the others will be "disabled." One goal of training the model is to raise the chances of the capsules that match the real mood class and lower the chances of the others. It is most similar between the reproduced capsule representation that matches the real class and the instance representation. For the other capsules, it is least similar. During testing, the capsule with the highest chance instance should be turned on and all the others should be turned off. The crop forecast class for that instance is the output of this capsule.

3.3.2 BiLSTM

There are problems with disappearing and exploding gradients in recursive neural networks. Also, RNN can't show changes in opinion over time. LSTM was made to handle long-term temporal relationships so that these problems could be fixed. RNN is what LSTM is made of. It fixes the gradient issues by adding some gates to the cell circuits. Instead of just input and output gates, LSTM has a memory line and gates like "Input," "Forget," and "Output." The vector that the LSTM layer makes by basically looking at the output from the cell before this one and the input from this one. In the end, the output of the last hidden layer is used to guess the mood.

To feed the BiLSTM layer with real-valued packed vectors, the embedded sequence is converted in the shown model. Next, we create two series vectors that are polar opposites to one another. Input, output, h_o forget, and x_o memory units are the four components of an LSTM cell, as we have already discussed. You can find the formulas for the computations in (4) through (9).

$$b_o = \rho(S_b c_o + E_b f_{o-1} + i_b) \tag{4}$$

$$h_o = \rho(S_h c_o + E_h f_{o-1} + i_h) \tag{5}$$

$$t_o = \rho(S_t c_o + E_t f_{o-1} + i_t) \tag{6}$$

$$e_o = tanf(S_e c_o + E_e f_{o-1} + i_e) \tag{7}$$

$$x_o = b_o * e_o + h_o * x_{o-1} \tag{8}$$

$$f_o = t_o * tanf(x_o) \tag{9}$$

The input at time o is shown by c_o, the sigmoid activation function is shown by ρ, and the elements' corresponding multiplication is shown by *. Also, c_o is used to lessen the problem of gradient loss or explosion, which helps LSTM find longer information dependence. The h_o forget gate resets the memory unit, and it and t_o show the input and output gates that are used to control the input and output of the memory unit. LSTM, on the other hand, can't encode information both forward and backward. For a certain amount of time, the bidirectional LSTM network is used to handle the features both forward and backward. Both forward and backward LSTM are combined with BiLSTM, which gives more background information and makes learning better and faster [24]. Figure 1 shows the shape of a biLSTM. Here, BiLSTM is used to look at the output of the capsule layer in order to improve how well the network features fit and how well they can generalise to a new set of data. Both hidden layers in BiLSTM are trained on the same set of inputs. One is trained on the original input series, and the other is trained on the inverted copy of the same input series. The goal is to link both hidden layers that go in opposite ways to the same output. In the new framework, an embedded sequence is turned into a set of real-valued dense vectors. Next, a one-dimensional convolutional layer is used to create a series of feature vectors that are too small, as shown in (5).

$$FV = (hk_1, hk_2, \ldots., hk_O) \tag{5}$$

The created vector is then sent to a BiLSTM layer, which makes two reverse sequence vectors shown in (6) and (7).

$$H_f = (\overrightarrow{f_1}, \overrightarrow{f_2}, \ldots., \overrightarrow{f_O}) \tag{6}$$

$$H_r = (\overleftarrow{r_1}, \overleftarrow{r_2}, .., \overleftarrow{r_O}) \tag{7}$$

Where H_f stands for the forward vector and H_r for the reverse vector.

3.3.3 Structure of Capsule

When training a BiLSTM, the input vectors for the capsule layer are generated by averaging the hidden vectors. The three elements that make up each capsule all work together. The visual illustration element uses the attention mechanism to create the capsule's case representation CV. The Likelihood component then estimates the active capsule instance's ni_b using an activation function such as sigmoid. Finally, the instance restoration representation is determined by the Restoration section by multiplying the active state chance ni_b by the capsule vector representation $x_{k,b}$. Rumour has it that in order to build an internal capsule model, the BiLSTM hidden vectors pass through the attention layer. It determines a word's significance for the current prediction task. In capsule, we employ a one-parameter method to increase attention:

$$u_{o,b} = f_o s_{y,b} \tag{8}$$

$$k_{x,b} = \sum_{d=1}^{P_w} y_o, b f_o \tag{9}$$

In tens and eleventh, f_o stands for the word representation at point o (RNN hidden vector), together with $s_{y,b}$ being the ith capsule's attention layer parameter. By multiplying the representations by the weight matrix, you may obtain the attention importance score ($u_{o,b}$) for every location. Words that compress the modulus of the input vectors have a normal chance distribution because of the compression function. You can use one module with any other. The property is represented by a capsule, and the instance's state is matched by the state of the capsule. Therefore, when the capsule's mood is identical to the first case, the probability module, which is based on capsule representation, is at its maximum. The state probability of the capsule representation is the basis of the rebuilding module. In other words, while it's in its "active" state, it can support representing an input instance.

4. Result and Discussion

A very important part of farming is the soil, which gives plants the nutrients they need to grow. There are different kinds of dirt, and each type has its own properties. Different kinds of plants grow on these different plots of land. Certain crops do better in certain types of soil, so we need to know what those kinds are and how they work.

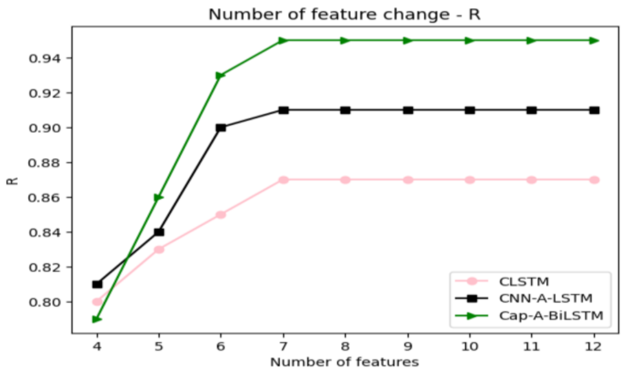

Fig. 1. The performance of models with k features and all of those features being offered for choice.

With the feature selection method, we let all features be chosen and then compared the outcomes from K = 4 to K = 12. It is clear that as K went up, so did the R value, but the rise in R levels stopped around K = 7 (Figure 1).

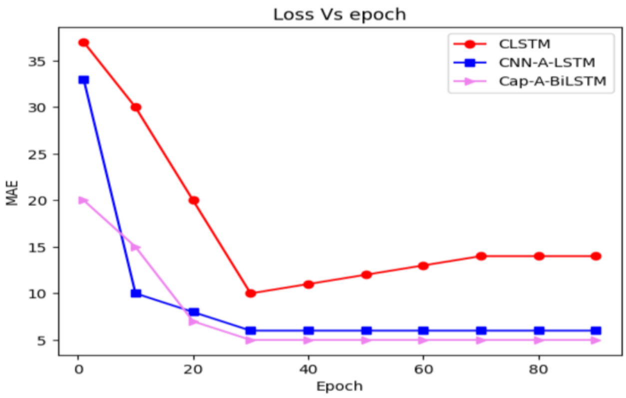

Fig. 2. Loss Vs Epoch of the Models

When the six deep neural networks are being trained, Fig. 2 shows the step where they come together. In the early stages of training for the suggested methods, the MAE figure losses drop sharply. They will stay close to their optimal values for the next iterations.

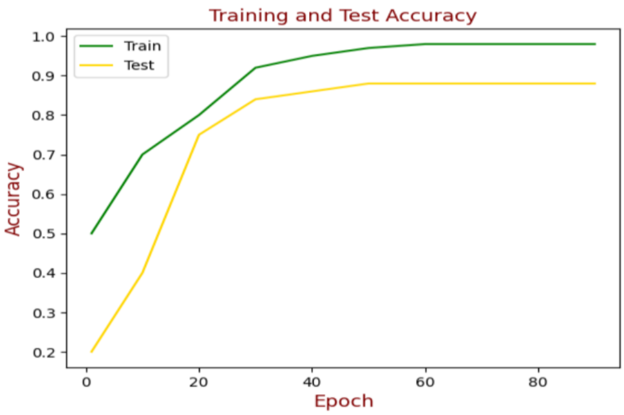

Fig. 3. Model Accuracy

In its first year, the plant is in the stage of sterilization and healthy hybridization. It has also not been exposed to a lot of chemicals and is growing in the right weather. After that, when the plant is in its later stages of growth and is exposed to more pesticides, different weather conditions, and bad farming practices, it will be at its most completely unhealthy state. 70% of the data was used for training and 30% for testing, and the number of batches needed for the model was found. Figure 3 shows how the accuracy rate was found.

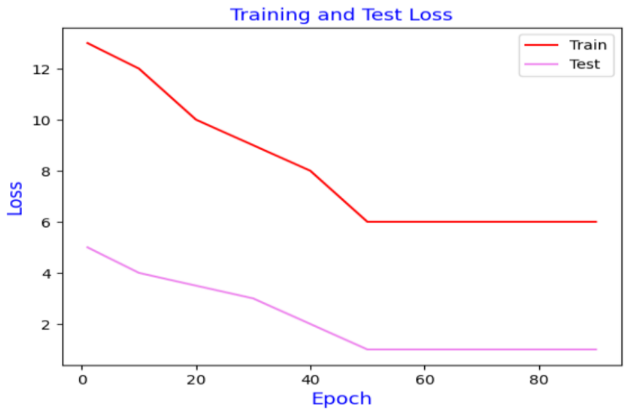

Fig. 4. Model Loss

As shown in Figure 4, the loss rate is almost 1%, which is the same as the accuracy rate.

5. Conclusion

Many nations' economies revolve around agriculture, and soil is essential to this sector. Soil types vary, and each sort is best suited to growing certain crops due to its unique characteristics. Modern research in this area has made use of a wide variety of models and techniques for increasing agricultural yields. Building a model that tells farmers which crops do best in various soil types is, hence, the primary goal of this approach. Plant recommendations are based on soil type or series utilizing machine learning algorithms implemented in this system. Just by inputting the soil type, the programme is able to recommend crops that will thrive in that particular environment. Different classifiers are employed in this case, and the model proposes the harvest based on their results. In this model, preprocessing, feature selection, and training the model are all used. This is where SG Denoising comes in. To choose features, CFS and MIFS are used. The last step is to use Cap-A-BiLSTM to train the model. The suggested method is weighed against two well-known ones, namely CLSTM and CNN-A-LSTM. The suggested method works well and gets results that are accurate about 98% of the time.

References

[1] H. Khan, A. A. Farooque, B. Acharya, F. Abbas, T. J. Esau, and Q. U. Zaman, “Delineation of management zones for site-specific information about soil fertility characteristics through proximal sensing of potato fields,” *Agronomy*, vol. 10, no. 12, 2020, doi: 10.3390/agronomy10121854.

[2] S. D. Roberton, J. McL. Bennett, C. R. Lobsey, and T. F. A. Bishop, “Assessing the sensitivity of site-specific lime and gypsum recommendations to soil sampling techniques and spatial density of data collection in Australian agriculture: A pedometric approach,” *Agronomy*, vol. 10, no. 11, pp. 0–10, 2020, doi: 10.3390/agronomy10111676.

[3] R. A. Viscarra Rossel, D. J. J. Walvoort, A. B. McBratney, L. J. Janik, and J. O. Skjemstad, “Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties,” *Geoderma*, vol. 131, no. 1–2, pp. 59–75, 2006, doi: 10.1016/j.geoderma.2005.03.007.

[4] M. N. Kathiravan, S. Bande, S. Tiwari, A. K. Koshariya, M. M, and D. Suganthi, “Soil Analysis and Classification to Improve the Crop Yield in Agriculture Using Fast Graph - CRNN Approach,” in *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Oct. 2023, no. Icssas, pp. 98–104, doi: 10.1109/ICSSAS57918.2023.10331819.

[5] A. R. Arunarani, “Crop Yield Prediction using Spatio Temporal CNN and Multimodal Remote Sensing,” *2023 2nd Int. Conf. Edge Comput. Appl.*, no. Icecaa, pp. 1042–1048, 2023, doi: 10.1109/ICECAA58104.2023.10212267.

[6] Thangavel, S., & Selvaraj, S. (2023). Machine Learning Model and Cuckoo Search in a modular system to identify Alzheimer's disease from MRI scan images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(5), 1753-1761.

[7] Saravanakumar, S., & Saravanan, T. (2023). Secure personal authentication in fog devices via multimodal rank-level fusion. Concurrency and Computation: Practice and Experience, 35(10), e7673.

[8] A. Nihar, N. R. Patel, and A. Danodia, "Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India," *J. Indian Soc. Remote Sens.*, vol. 50, no. 8, pp. 1519–1530, 2022, doi: 10.1007/s12524-022-01549-0.

[9] A. Crane-Droesch, "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture," *Environ. Res. Lett.*, vol. 13, no. 11, 2018, doi: 10.1088/1748-9326/aae159.

[10] S. Iniyan, V. Akhil Varma, and C. Teja Naidu, "Groundnut Crop Yield Prediction Using Machine Learning Techniques," *Adv. Eng. Softw.*, vol. 175, no. July 2018, 2023, doi: 10.1016/j.advengsoft.2022.103326.

[11] Saravanakumar, S., & Thangaraj, P. (2019). A computer aided diagnosis system for identifying Alzheimer's from MRI scan using improved Adaboost. Journal of medical systems, 43(3), 76.

[12] A. A. Farooque, Q. U. Zaman, A. W. Schumann, A. Madani, and D. C. Percival, "Response of wild blueberry yield to spatial variability of soil properties," *Soil Sci.*, vol. 177, no. 1, pp. 56–68, 2012, doi: 10.1097/SS.0b013e3182376ed6.

[13] N. R. Kitchen, S. T. Drummond, E. D. Lund, K. A. Sudduth, and G. W. Buchleiter, "Soil electrical conductivity and topography related to yield for three contrasting soil-crop systems," *Agron. J.*, vol. 95, no. 3, pp. 483–495, 2003, doi: 10.2134/agronj2003.4830.

[14] Kumaresan, T., Saravanakumar, S., & Balamurugan, R. (2019). Visual and textual features based email spam classification using S-Cuckoo search and hybrid kernel support vector machine. Cluster Computing, 22(Suppl 1), 33-46..

[15] A. A. Farooque *et al.*, "Evaluation of DualEM-II sensor for soil moisture content estimation in the potato fields of Atlantic Canada," *Plant, Soil Environ.*, vol. 65, no. 6, pp. 290–297, 2019, doi: 10.17221/72/2019-PSE.

[16] Saravanakumar, S. (2020). Certain analysis of authentic user behavioral and opinion pattern mining using classification techniques. Solid State Technology, 63(6), 9220-9234..

[17] D. Heckmann, U. Schlüter, and A. P. M. Weber, "Machine Learning Techniques for Predicting Crop Photosynthetic Capacity from Leaf Reflectance Spectra," *Mol. Plant*, vol. 10, no. 6, pp. 878–890, 2017, doi: 10.1016/j.molp.2017.04.009.

[18] S. Mamun, S. K. Shuvessa, A. Al Ryan, N. Nabi, and M. S. Ahamed, "JuteBangla: A Comparative Study on Jute Yield Prediction using Supervised Machine Learning Approach based on Bangladesh Perspective," *2022 13th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2022*, no. May 2023, 2022, doi: 10.1109/ICCCNT54827.2022.9984311.

[19] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *J. Big Data*, vol. 4, no. 1, 2017, doi: 10.1186/s40537-017-0077-4.

[20] V. A. Bharadi, P. P. Abhyankar, R. S. Patil, S. S. Patade, T. U. Nate, and A. M. Joshi, "Analysis and Prediction in Agricultural Data Using Data Mining Techniques," *Int. J. Res. Sci.*, no. 7, pp. 386–393, 2017.

[21] L. J. Armstrong, "Predicting Rice Crop Yield Using Bayesian Networks," *Conf. Adv. Comput. Commun. Informatics*, pp. 795–799, 2016.

[22] R. Biswas and A. Sarkar, *'Omics' Tools in Soil Microbiology: The State of the Art*, no. March 2018. 2018.

[23] S. M. Prober *et al.*, "Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands

worldwide," *Ecol. Lett.*, vol. 18, no. 1, pp. 85–95, 2015, doi: 10.1111/ele.12381.

[24] D. C. Schlatter, M. G. Bakker, J. M. Bradeen, and L. L. Kinkel, "Plant community richness and microbial interactions structure bacterial communities in soil," *Ecology*, vol. 96, no. 1, pp. 134–142, 2015, doi: 10.1890/13-1648.1.

[25] T. S. De Carvalho *et al.*, "Land use intensifcation in the humid tropics increased both alpha and beta diversity of soil bacteria," *Ecology*, vol. 97, no. 10, pp. 2760–2771, 2016, doi: 10.1002/ecy.1513.