ISSN 0315-3681 Volume 122 (2), 2025

GREENHOUSE YIELD PREDICTION USING FEATURE SELECTION AND ENHANCED ARTIFICIAL NEURAL NETWORK ALGORITHM

[1] Dr. P. Suresh Babu, [2] Ms. P. Karthika*

[1] Associate Professor & Head, [2] Research Scholar

[1,2] Department of Computer Science, Bharathidasan College of Arts and Science, Erode

E-Mail: [1] ptsuresh77@gmail.com, [2] karthika3141@gmail.com

Abstract

Greenhouse yield prediction is a crucial aspect of modern agriculture, aiming to optimize production and resource management in controlled environments. Accurate yield prediction in greenhouse agriculture is paramount for optimizing resource allocation and maximizing productivity, yet the complex interplay of numerous environmental and operational factors poses a significant challenge. In this work, Improved Chicken Swarm Optimization (ICSO) and Enhanced Artificial Neural Network (EANN) algorithm is proposed. The main steps of this research includes pre-processing, feature selections, and classifications for greenhouse yield prediction. Initially min-max normalization algorithm is proposed to improve the quality of the given dataset. Then, ICSO algorithm is introduced for feature selection which selects more relevant and significant features from the given dataset. These ICSO variants, by leveraging their enhanced search capabilities and adaptability, provide robust and effective solutions for optimizing greenhouse management, resource allocation, and yield prediction. It generates best fitness values based on the higher accurate features. Finally, EANN is applied to perform the yield prediction in the given greenhouse dataset. EANN is focused to effectively manage the complexities of greenhouse environments and optimize crop yields via hidden neurons. Also, it is used for optimizing environmental controls, detecting diseases, and managing resources efficiently, ultimately leading to increased productivity and sustainable greenhouse operations. The training and testing process is conducted and it is used to provide more accurate results. The experimental results prove that the proposed APSO-EANN algorithm provides better greenhouse yield results in terms of higher accuracy, precision and RMSE values than the existing algorithms

Key words: Greenhouse yield prediction, Improved Chicken Swarm Optimization (ICSO), feature selection, Enhanced Artificial Neural Network (EANN) algorithm

1. Introduction

In recent decades, intensification of agricultural production has been achieved by increasing input application, especially fertilizer, to strengthen the agri-food system and feed the increasing global. Achieving high crop yield from the conventional production system requires multiple tillage operations for land preparation and planting as well as applications of high amounts of fertilizers, irrigation, and pesticides. These can result in increased energy inputs and consequently increased greenhouse gases (GHG) emissions. In intensive cereal production systems, energy use is highest from fertilizers followed by fuel use for irrigation and machinery required for tillage and other farm operations. Increased fertilizer application is necessary to increase cereal yields to feed the increasing global population but increased fertilizer use can also emit substantial amounts of GHGs and contribute to global warming [1] [2]. To develop optimum fertilizer recommendations to crops and simultaneously reduce global warming, fertilizer rates are often determined for site-, soil- and crop-specific situations. However, there is an alternative and more robust approach to fertilizer recommendation, known as 'yield-scaled emissions (YSE)' or 'Greenhouse Gas Emission Intensity (GHGI)'. This approach considers the trade-off among yield, economics, energy, and GHGs emissions, and has now been used by several researchers for achieving the sustainable intensification of agriculture.

Crop yield forecasting in greenhouses plays an important role in farming planning and management in greenhouses, and optimally controlling environmental parameters guarantees the maximum crop yield. Cultivators and farmers can utilize yield prediction in greenhouses to make knowledgeable management and financial decisions. However, it is an extremely challenging task [3]. There are many factors that have an influence on crop yield in a greenhouse, such as radiations, carbon dioxide concentrations, temperature, quality of crop seeds, soil quality and fertilization, and disease occurrences. It is not straightforward to construct an explicit model to reflect the relationship between such a variety of factors and crop yield.

The greenhouse microclimate system is a strongly coupled nonlinear system [4]. However, alterations to the greenhouse environment are typically influenced by external climatic conditions, crop growth, and the structure and materials of the greenhouse itself. Consequently, the multitude of heat and material exchange processes that affect changes in the greenhouse environment is inherently complex. However, energy-saving greenhouse climate regulation usually requires the accurate prediction of the climate inside the greenhouse, therefore improving the simulation performance of the greenhouse climate model becomes a critical issue.

Fig 1 Example of greenhouse crop yield

With ongoing development, deep learning network models have also been applied to predicting greenhouse environments [5]. It employed deep neural network technology to enhance greenhouse modeling and control tasks. Optimization algorithms to adjust the parameters of neural network models, thereby ensuring the predictive performance of these models. However, traditional neural networks have limitations in their learning capabilities and struggle to capture the complex patterns of changes in greenhouse environments. In recent years, a new machine learning modeling approach that combines nonlinear time series models with neural networks has gained widespread attention [6]. The modeling methods Artificial Neural Networks (ANN), Nonlinear Autoregressive Exogenous Models (NARX), and Recurrent Neural Networks with Long Short-Term Memory (LSTM) are applied for greenhouse environmental changes [7]. The results indicated that the LSTM neural network exhibits superior approximation performance for these changes and effectively addresses the precise modeling of greenhouse environmental time series.

The main objective of this research work is the greenhouse crop yield prediction using ICSO-EANN algorithm. There are several research and methodologies introduced but the crop yield classification accuracy is not achieved significantly. The existing approaches has drawback with efficient feature selection algorithm results. The main contribution of this research is preprocessing, feature selection and greenhouse crop yield prediction. The proposed method is used to provide more accurate classification results using effective algorithms for the given greenhouse dataset.

The remaining portions of the essay are structured as follows: In Section 2, there is a short discussion of some of the studies in the field of feature selection, and greenhouse crop yield prediction. Section 3 provides specifics on the suggested technique for the ICSO-EANN scheme. Section 4 contains a description of the experimental performance analysis findings. Finally, Section 5 summarizes the results.

2. Related work

In [8], Chen et al (2022) focused to study the design of the multi-energy supply system based on the adaptive improved genetic algorithm for the intelligent control system of agricultural greenhouses. The related concepts of the improved genetic algorithm were introduced, and a framework was designed for the agricultural greenhouse control system. In order to verify the feasibility of the developed software and hardware, a greenhouse experiment platform was designed and improved in the laboratory, and the system designed in this paper was debugged and tested. The experimental results show that through intelligent control of agricultural greenhouses, the deviation of air temperature is less than 0.5 °C and the deviation of air humidity is less than 1% RH when stable. The value of carbon dioxide concentration after being basically stable has small fluctuations around the stable value, and the maximum relative fluctuation is less than 2.5%.

In [9], Seyedmohammadi et al (2023) aimed to predict yield and sustainably manage the use of natural resources such as soil and water, we modelled the effect of soil properties by classification and regression tree, *k*-nearest neighbors, support vector machines and developed a new hybrid model of support vector machines and the firefly meta-heuristic algorithm. We sampled soils from 124 pistachio orchards in Iran and analyzed them for a range of parameters. According to the results, *k*-nearest neighbors, classification and regression tree and support vector machines algorithms could explain 83, 84 and 88% of the variation of pistachio yield, respectively, but improved to 94% in the hybrid model because it was more able to efficiently capture non-linear relationships. Soil available phosphorus was the most important determinant of pistachio yield, with soil salinity, exchangeable sodium percentage, potassium, gypsum, calcium carbonate and gravel ranked in order of decreasing importance. These outputs can help planners and farmers to better manage soil properties to increase pistachio yield and sustainable production

In [10], García-Vázquez et al (2023) presented Linear Regression (LR) and Support Vector Regression (SVR) to predict the internal temperature of a greenhouse. A meteorological station is installed in the greenhouse to collect internal data (temperature, humidity, and dew point) and external data (temperature, humidity, and solar radiation). The data comprises a one year, and is divided into seasons for better analysis and modelling of the internal temperature. The study involves sixteen experiments corresponding to the four models and the four seasons and evaluating the models' performance using R2, RMSE, MAE, and MAPE metrics, considering an acceptability interval of ±2 °C. The results show that LR models had difficulty maintaining the acceptability interval, while the SVR models adapted to temperature outliers, presenting the highest forecast accuracy among the algorithms

3. Proposed methodology

In this research, ICSO-EANN algorithm is proposed to improve the greenhouse crop yield prediction results for the given greenhouse dataset. The proposed system contains main phases are such as preprocessing, feature selection and greenhouse crop yield prediction. Overall block diagram of the proposed system is shown in Fig 2

3.1 Dataset collection

Dataset is collected to predict the future crop yield based on historical yields and greenhouse environmental parameters (e.g., CO_2 concentration, temperature, humidity, radiation, etc.) information

3.2 Data pre-processing using Min max normalization

In this work, data pre-processing is done by using min max normalization algorithm. The algorithm aims to increase greenhouse crop yield prediction accuracy. Normalization is accomplished once we remove the values lacking from the dataset. Because input might have scale versions that cause inaccurate outcomes, normalizing the records is required to avoid these problems. The normalization system includes converting numerical values into a brand-new range with a mathematical characteristic. Minimal standardization is one of the maximum mutual conducts to regularize facts. The ethics within the dataset are standardized within the assumed variety of minimum and maximum values from the dataset

Min-max scaling, also referred to as normalizing, is the process of rescaling numerical properties to a certain range, typically between 0 and 1, is a data preparation technique [11]. For machine learning algorithms that require inputs to fall within a certain range, this procedure is especially helpful when the initial feature values have different scales and need to be standardized.

$$d' = \frac{[d-\min(p)]*[new_{\max(p)}-new\_\min(p)]}{[\max(p)-\min(p)]} + new\_\min(p) \quad (1)$$

where min(p) = minimum value of attribute, max(p) = maximum value of attribute

In case min-max normalization maps a value d of P to d′ in the range [0,1], so put new_min(p) =0 and new_max(p) =1 in the above equation (2). Now we get the simplified formula of min-max normalization

$$d' = \frac{d-\min(p)}{\max(p)-\min(p)} \quad (2)$$

d' is new value of normalized result. Min max normalization preserves the relationship among the original data values

A crucial step in preparing data for analysis and modeling is filtering null values to guarantee the quality and integrity of the dataset. This research aims to remove characteristics that have a significant percentage of null values, i.e., 80% or more of null values. The reasoning behind excluding qualities that have a substantial percentage of null values is due to the possible distortion that these attributes may cause in the analysis. If null values are not handled appropriately, it could affect statistical measurements, impair model performance, and produce untrustworthy findings. By removing characteristics that have a high frequency of null values, can reduce the possibility of generating incorrect conclusions and guarantee that the information employed for modeling is more reliable and representative. Furthermore, by making managing missing data less complicated, deleting characteristics with a large percentage of null values makes the remaining data preparation procedures simpler. It simplifies the pipeline for analysis and enables to concentrate on significant traits that make a significant contribution to the current greenhouse crop yield dataset prediction

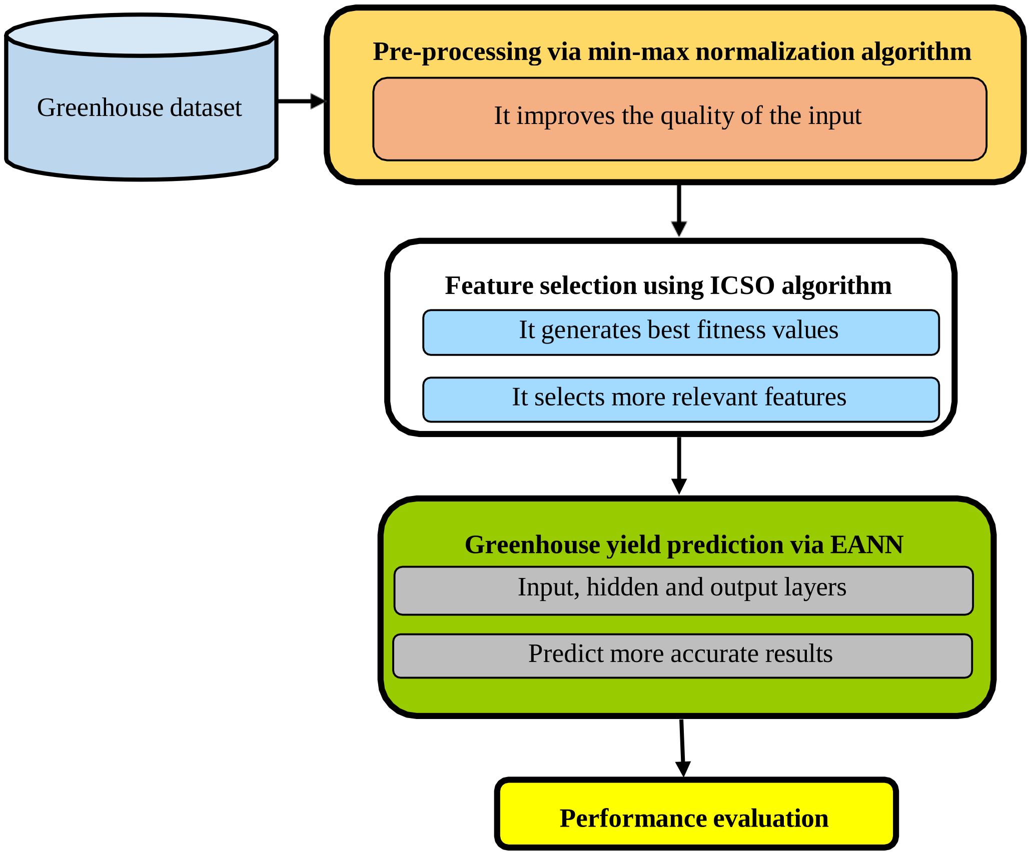

Fig 2 Overall block diagram of the proposed system

3.3 Feature selection using Improved Chicken Swarm Optimization (ICSO) algorithm

In this work, feature selection is done by using ICSO algorithm to improve greenhouse crop yield dataset prediction accuracy. The algorithm imitates hierarchies of chicken swarms and individual chickens' behaviours. The hierarchy of a swarm of chickens divides into numerous groups, each consisting of a rooster, several hens, and chicks. distinct kinds of chickens are subject to distinct laws of motion. Hierarchical order plays a key role in hens' social interactions. The most powerful chickens in a flock will subjugate the weaker ones. Submissive hens stand outside the groupings, while more dominant hens kept closer to the roosters. The nature of the DCO algorithm is seen in Fig 3.

Fig 3 Nature of CSO algorithm

Chickens Movements

Rooster Movement: Better-fit roosters have a larger range of locations where they can look for food.

Hen movement: Hens in a flock follow the roosters to locate food. The other chickens would also put restrictions on them, but they would heedlessly pilfer the delicious food the other birds had discovered. When it came to the struggle for food, the stronger chickens would win out over the weaker ones.

Chick movement: In order to find nourishment, the chicks roam about their mother.

The regulations that follow explain the actions of the chickens and provide the foundation of the mathematical model of CSO utilized in [12]:

1) chicken swarms have different groups. There are dominant roosters in groups, proceeded by several hens and chicks.

2) The roosters which act as group leaders have the maximum fitnesses, while chicks have least fitnesses, illuminating the organizational structure of the swarm. The hens would be the other group.

3) The mother-child relationship and the group's power dynamics won't change. Only a few time steps separate updates for these states.

4) The virtual chickens in the swarm are separated into n groups as follows: R_n, C_n, $H_n$and M_n which are counts of roosters, chicks, hens, and the mother hens, correspondingly. The positions of each individual in a D-dimensional space serve as a representation of

$$x_{i,j}(i \in [1,\ldots,N], j \in [1,\ldots,D]). \quad (3)$$

Greater fitness values are given preference over lower fitness standards in terms of feeding roosters. The fact that roosters with higher fitness values can forage in a wider range of environments than roosters with lower fitness values makes it easy to replicate the situation. The use of the Gaussian distribution helps to greatly improve the energy consumption because the normal chicken swarm has a problem with them. The ICSO may then be created as shown below.

$$x_{i,j}^{t+1} = x_{i,j}^{t} * (1 + Randn(0, \sigma^2)) \quad (4)$$

$$\sigma^2 = \begin{cases} 1, & if\ f_i \leq f_k \quad k\epsilon[1, N], k \neq i \\ \exp\left(\frac{(f_k - f_i)}{|f_i| + s}\right), & \text{otherwise} \end{cases} \quad (5)$$

Where Randn (0, σ^2) are Gaussian distributions with means 0 and standard deviationsσ^2, The lowest constant, ε, is employed to avoid mistakes in zero-division calculations. The term "f" denotes matching x's fitness levels, and the indices of the roosters are randomly picked among groups of roosters.

The roosters lead the hens in their flock on a food hunt. Additionally, they would clumsily steal the delectable food that other had found even when restricted by the other chickens. The superior hens would profit from food competition in contrast to more submissive birds. These issues can be formally expressed as follows.

$$x_{i,j}^{t+1} = x_{i,j}^{t} + S_1 * rand * (x_{r1,j}^{t} - x_{i,j}^{t}) + S_2 * rand * (x_{r2,j}^{t} - x_{i,j}^{t}) \quad (6)$$

$$S_1 = \exp\left((f_i - f_{r1})/(abs(f_i) + \varepsilon\right) \quad (7)$$

$$S_2 = \exp\left((f_{r2} - f_i)\right) \quad (8)$$

rand is a consistent random number between [0, 1], where. While $r2\epsilon[1, .. N]$ is an index of the chicken, which is randomly chosen from the swarm $r1 \neq r2$, $r1\epsilon[1, .. N]$ is a representation of the rooster, the ith hen's groupmate.

Of course, $f_i > f_{r1}$, $f_i > f_{r2}$,, therefore S2 <1< S1. Assuming S1=0, the ith hen would feed for food before the other chickens. Greater fitness value and lower S2 are seen between two chicks whose locations are more dissimilar from one another. Thus, there would be a reduced likelihood of the chickens stealing food found by other birds. As a result of intragroup contests, S1's formula form differs from S2's. To make things simpler, the contests amongst hens in a set are represented by the chickens' fitness values in relation to the rooster's fitness value. When S2 is equal to zero Within its own domain, the i-th hen would look for food. For that group, the rooster has a unique fitness value [14]. Thus, the more S1 resembles 1 and the closer its position is to that of its groupmate rooster, the lower the i-th hen's fitness value. Consequently, the stronger hens are more likely than the weaker birds to take the food.

Young birds move around their mother to discover nourishment which is presented as follows

$$x_{i,j}^{t+1} = x_{i,j}^{t} + FL * (x_{m,j}^{t} - x_{i,j}^{t}) \quad (9)$$

Where $x_{m,j}^{t}$are positions of ith chick's mothers ($m\epsilon[1, N]$). $FL(FL\epsilon(0,2))$ are parameters, suggests that chicks might go hunting with their mothers.

There will be a maximum number of parameters. The amount of parameters that require optimization is indicated by the total number of parameters (TNP). Temperature (Temp), Humidity (Humd), Carbon dioxide (CO2), and Sunlight (S) are the four variables we are optimizing. The upper bound of the parameters i is represented by UpBdi, Where i=1,2,3,...,TNP, with TNP representing the total number of parameters requiring optimization. The maximum values for Humidity (Humd) is 80%, Temperature (Temp) is 24°C, Sunlight intensity (S) is 400 μmdi/ m2/sec and Carbon dioxide (CO2) is 1000 ppm. The lower bound of the parameters i is represented by LowBdi, Where i=1,2,3,...,TNP, with TNP representing the total number of parameters requiring optimization. The minimum values for Humidity (Humd) is 40%, Temperature (Temp) is 18 °C, Sunlight intensity (S) is 300 μmdi/ m2/sec and Carbon dioxide (CO2) is 400 ppm. "Range" refers to the difference of the parameters' upper and lower boundaries as given by UpBdi—LowBdi. There are four different ranges: 40, 600, 100, and 6 for humidity, carbon dioxide, sunlight, and temperature respectively.

Algorithm 1: ICSO for optimal feature selection

Input: Features from greenhouse dataset

Objective function: better accuracy

Output: Optimal feature selection

1. Initialize the parameters such as R_n, H_nC_n, and M_n(greenhouse dataset)
2. Estimates N chickens' fitness values, t=0; (higher classifier accuracy)
3. While $t < Maximum$ iteration do
4. If t%G==0 then
5. Rank feature's fitness values and create hierarchal orders in swarms
6. Optimize the parameters such as Temperature (Temp), Humidity (Humd), Carbon dioxide (CO2), and Sunlight (S)
7. Separate swarms into various groups and ascertain how each groups' mothers and chicks are related to one another.
8. End
9. For i=1 to N nodes (higher classifier accuracy) do
10. If i==rooster then
11. Update the solution using (4) for each feature i that is to be assigned to a crop yield growth do
12. End if
13. If i==hen then
14. Update the solution using (6) for each feature i that is to be assigned to a crop yield growth do
15. End if
16. If i==chick
17. Update the solution using (9) for each feature i that is to be assigned to a crop yield growth
18. End if
19. Estimate the novel solution
20. Update it if the newly created solution is superior to the prior one.
21. End
22. End
23. Return the best features for the dataset

Here, utilizing the greenhouse dataset feature arbitrary positions as a means of allocating and identifying the chicks. The highest fitness chicken is updated to enhance the key node of the Algorithm 1-described for the given dataset. Using equation (8), determine the best hens (features) with global best value and users searches for fitness values by computing best values for all chickens. It provides energy efficiency, convergence speed, and maintaining optimal conditions for plant growth

Plants comfort

Comfort value is a metric used in optimization algorithms to measure how comfortable the Plant is with the algorithm's results. It is a measure of Plant's satisfaction with the solution presented by the algorithm and is used to determine whether the algorithm is providing satisfactory results or not. Comfort value is usually calculated by analyzing the Plant Preferred Environment, such as preferences, and comparing it to the algorithm's output. If the output matches the Plant Preferred Environment, then the comfort value is high. If the output does not match the Plant Preferred Environment, then the comfort value is low. To calculate the Comfort Index, we used the following formula

$$CI = prr1[1-(\frac{err1}{T_u})^2] + prr2[1-(\frac{err2}{C_u})^2] + prr3[1-(\frac{err3}{S_u})^2] + prr4[1-(\frac{err4}{H_u})^2] \quad (10)$$

CI denotes the Plant comfort level index. The parameters set by the Plant Preferred Environment, namely err1, err2, err3, and err4, correspond to the four respective parameters, and the Environmental parameters are given by err1, err2, err3, and err4. CI can have a maximum value of 1. Cu is the Plant Preferred concentration of carbon dioxide, Su is the Plant Preferred amount of sunlight, Hu is the Plant Preferred humidity, and Tu is the Plant Preferred temperature

3.4 Greenhouse yield prediction using Enhanced Artificial Neural Network (EANN) algorithm

In this work, greenhouse yield prediction is done by using EANN algorithm for the given dataset more effectively. ANN learns to gather knowledge and it has three phases such as input layer, hidden layer, and output layer. The input layer collects input data attributes and processes them to generate 'n' inputs. These processes adhere to a set of weights. Weights are the information used to solve neural network problems [17]. After some beneficial hidden extraction, the hidden information is taken from the input layer and sent to the output layer. EANN is utilised in this instance to classify greenhouse dataset. EANN is used to train the selected feature dataset, and the features are categorised by state during testing. MLP (Multilayer Perceptron) and ANN are combined to create the EANN. The ANN architecture is depicted in Fig 4.

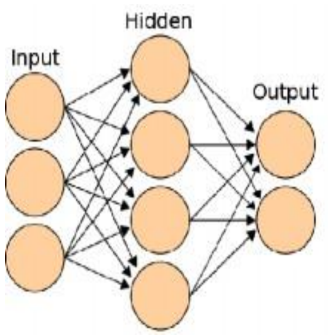

Fig 4 Architecture of ANN

Input Layer - The network's input layer carries the chosen features of the given greenhouse database. At first, this material seems pretty undeveloped.

Hidden Layer - Primary functions of these layer are to transform raw inputs into decidable outputs where EANN architecture may encompass multiple hidden layers.

Output Layer - The output layer receives data from the hidden layer and processes it to provide the desired outcomes (better classifier accuracy and faster execution).

The MLP FNN (Feedforward Neural Network) architecture, in which neurons are organised cascade-wise, is the most well-known FNN model. At least two layers make up MLP. In MLPs, there is no information transfer between the neurons that make up a layer; instead, inputs to neurons of i+1th layers are outputs of ith layers. Both counts of nodes in input layers and counts of nodes in output layers correspond to counts of features included in input vectors.

$$Y_n = f(\sum_{m=1}^{h} (w_{nm}, f(\sum_{l=1}^{i} v_{ml} X_l + \theta_{vm}) + \theta_{wm} \tag{11}$$

$$n = 1, \ldots\ldots, o$$

where Y_n stands for nth node's output layer, X_l implies inputs of lth nodes in input layers, w_{nm} represents connective weights between m nodes in hidden layers and n output layers, v_{ml} represents connective weights between nodes l in input layers and m stands for hidden layers and θ_{vm} and θ_{wm} represent bias terms or thresholds of transfer functions f of m nodes in hidden layers and n output layers.

In EANN, if the weighted total of the inputs is larger than a programmable cutoff value, also known as an activation function, the perceptron model transmits the output 1. Each neuron's output is the weighted total of its inputs, including its bias. The weights and input neuron, respectively, are denoted by "w" and "x."

$$\sum_{i=1}^{m} bias + (w^i x^i) \tag{12}$$

Activation function uses one of the specified functions is Sigmoid function

$$f(x) = sigmoid = \frac{1}{1+\exp(-x)} \tag{13}$$

The network weights are made up of the connection weights and bias terms of each neuron. The process of updating the network weights and determining the right weights and biases values is known as "neural network training," and it is thought to be the most efficient way to get the desired output from the input.

Algorithm 2: EANN for classification

Input: Selected features (given greenhouse databases) are used as input.
Output: Improved prediction performance

1. EANN procedure (input, neurons, repeat)
2. Make an input database.
3. Input←database with all possible combinations
4. Training via EANN
5. Do for input = 1 to end of input
6. Do for neurons = 1 to n
7. Do for repeat = 1 to n
8. Train EANN-storage ←save value with highest accuracy features
9. End for
10. End for
11. EANN-storage←save best prediction of EANN depending on inputs
12. End for
13. Return EANN-storage →Result with best classification of EANN for every feature combination

4. Experimental result

The experimental studies of the proposed ICSO-EANN based crop yield prediction approach are presented in this section. The dataset is collected from https://zenodo.org/records/6697044. It contains parameters are such as Timestamp & Environmental Factors, temperature, heat & radiation, gas & air properties, ventilation & screen properties, carbon & biomass, crop water & growth indicators, key variables for yield prediction.

4.1 Accuracy

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{14}$$

where, TP -True Positive, FN - False Negative, FP - False Positive and TN- True Negative

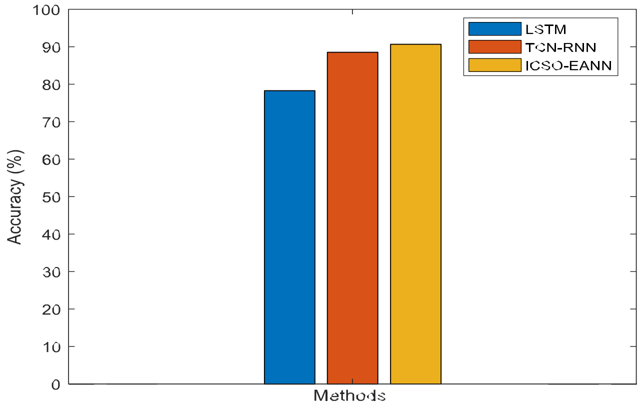

Fig 5 Accuracy

From the above Fig 5, it can be observed that the comparison metric is evaluated using existing and proposed methods in terms of accuracy. For x-axis the methods are taken and in y-axis the accuracy value is plotted. The proposed ICSO-EANN algorithm provides higher accuracy while the existing LSTM, TCN-RNN algorithms provide lower accuracy values for the given greenhouse yield dataset. Thus the result concluded that the proposed ICSO-EANN algorithm increase the greenhouse prediction accuracy for the given greenhouse dataset.

4.2 Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (15)$$

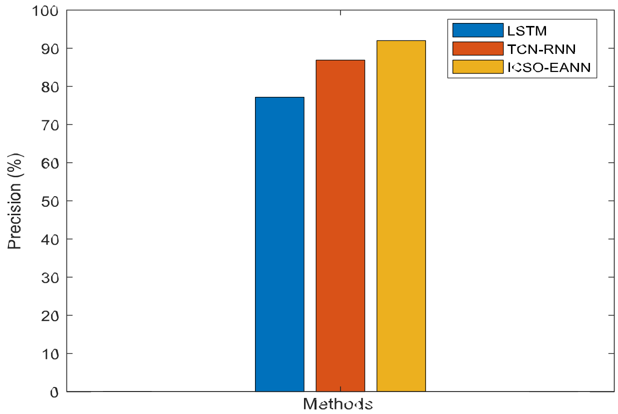

Fig 6 Precision

From the above Fig 6, it can be observed that the comparison metric is evaluated using existing methods in terms of precision. For x-axis the methods are taken and in y-axis the precision value is plotted. The proposed ICSO-EANN algorithm provides higher precision whereas the existing LSTM, TCN-RNN algorithms provide lower precision values for the given greenhouse dataset. Thus the result concluded that the ICSO-EANN algorithm increase the greenhouse prediction accuracy for the given dataset.

4.3 Root Mean Square Error (RMSE)

It measures the average difference between values predicted by a model and the actual values. It provides an estimation of how well the model is able to predict the target value (accuracy).

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}||y(i)-\hat{y}(i)||^2}{N}}x \quad (18)$$

where N is the number of data points, y(i) is the i-th measurement, and yˆ(i) is its corresponding prediction.

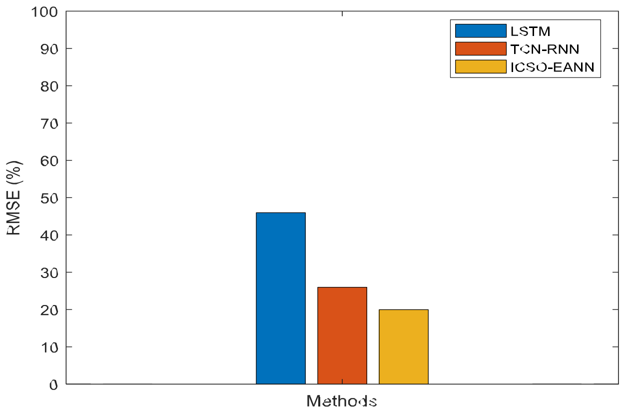

Fig 7 RMSE

From the above Fig 7, it can be observed that the comparison metric is evaluated using existing methods in terms of RMSE. For x-axis the methods are taken and in y-axis the RMSE value is plotted. The LSTM and TCN-RNN algorithms provide higher RMSE whereas the proposed ICSO-EANN algorithm provide lower RMSEs values for the given greenhouse dataset. Thus the result concluded that the ICSO-EANN algorithm increase the greenhouse prediction accuracy for the given dataset.

5. Conclusion

In this research work, ICSO-EANN algorithm is proposed to improve the greenhouse prediction accuracy results for the given dataset prominently. In this work, initially, min max normalization algorithm is used for pre-processing the dataset. Then, ICSO algorithm is applied for optimal feature selection which is used to select important and relevant feature from the greenhouse dataset. Finally, EANN classifier is proposed to greenhouse crop prediction. The result concludes that the proposed ICSO-EANN method provides higher accuracy, precision and lower RMSE values compare than the existing LSTM and TCN-RNN algorithms. In future research, optimization based deep learning algorithm can be developed for increasing the classifier accuracy.

References

1. Kocian, Alexander, et al. "Dynamic Bayesian network for crop growth prediction in greenhouses." *Computers and electronics in agriculture* 169 (2020): 105167.
2. Hu, Guoqing, and Fengqi You. "Model predictive control for greenhouse condition adjustment and crop production prediction." *Computer Aided Chemical Engineering*. Vol. 51. Elsevier, 2022. 1051-1056.
3. Yuan, Hongliang, Yuhui Liang, and Zhulin Li. "Development of autonomous navigation system based on neural network and visual servoing for row-crop tracking in vegetable greenhouses." *Smart Agricultural Technology* 9 (2024): 100572.
4. Rodríguez, Francisco, et al. "A multilayer control architecture for greenhouse crop production in agro-industrial districts: Conceptual framework, prospects and challenges." *Smart Agricultural Technology* (2024): 100657.
5. Sunil, G. C., et al. "Weed and crop species classification using computer vision and deep learning technologies in greenhouse conditions." *Journal of Agriculture and Food Research* 9 (2022): 100325.
6. Liu, Tan, Qingyun Yuan, and Yonggang Wang. "Hierarchical optimization control based on crop growth model for greenhouse light environment." *Computers and Electronics in Agriculture* 180 (2021): 105854.
7. Jung, Dae-Hyun, et al. "Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse." *Computers and Electronics in Agriculture* 173 (2020): 105402.
8. Chen, Qiuxia, and Xinghong Hu. "Design of intelligent control system for agricultural greenhouses based on adaptive improved genetic algorithm for multi-energy supply system." *Energy Reports* 8 (2022): 12126-12138
9. Seyedmohammadi, Javad, et al. "A new robust hybrid model based on support vector machine and firefly meta-heuristic algorithm to predict pistachio yields and select effective soil variables." *Ecological Informatics* 74 (2023): 102002

10. García-Vázquez, Fabián, et al. "Prediction of internal temperature in greenhouses using the supervised learning techniques: Linear and support vector regressions." *Applied Sciences* 13.14 (2023): 8531
11. Kappal, Sunil. "Data normalization using median median absolute deviation MMAD based Z-score for robust predictions vs. min–max normalization." *London Journal of Research in Science: Natural and Formal* 19.4 (2019): 39-44.
12. Ahmed, Khaled, Aboul Ella Hassanien, and Siddhartha Bhattacharyya. "A novel chaotic chicken swarm optimization algorithm for feature selection." *2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, 2017.
13. Wang, Haoran, Zhiyu Chen, and Gang Liu. "An improved chicken swarm optimization algorithm for feature selection." *International conference on wireless communications, networking and applications*. Singapore: Springer nature Singapore, 2021.
14. Pohan, S., B. Warsito, and S. Suryono. "Backpropagation artificial neural network for prediction plant seedling growth." *Journal of Physics: Conference Series*. Vol. 1524. No. 1. IOP Publishing, 2020.