Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

STRENUOUS GREY WOLF OPTIMIZATION-BASED FEED-FORWARD NEURAL NETWORK (SGWO-FFNN) FOR ENHANCED CROP YIELD PREDICTION

SHANMUGA PRIYA S[1], DR.M.SENGALIAPPAN[2]

[1]Research Scholar, Kovai Kalaimagal College of Arts & Science, Bharathiar University, Coimbatore

[2]Associate Professor & Head, Department of MCA, Nehru College of Management, Coimbatore

E-mail: [1]shanpriya.s@gmail.com, [2]cmsengs@gmail.com

ABSTRACT

One of the most profound changes to human life brought about by technological progress has been increased food production and other essentials. Machine learning techniques have contributed to this change by making our lives easier and bringing us closer by eliminating hunger and poverty. However, these techniques can potentially cause harm if not correctly applied. Machine learning-based crop yield predictions resulted in lower yields, as predicted yields can be as low as 20% of actual yields, and it is because of the poor performance of classification algorithms. The "Strenuous Grey Wolf Optimization-based Feed-Forward Neural Network (SGWO-FFNN)" to predict crop yield prediction is a machine learning model that combines the optimization algorithm called Grey Wolf Optimization (GWO) with a feed-forward neural network (FFNN). The GWO algorithm is a population-based optimization method that is inspired by the hunting behavior of grey wolves. The goal of the SGWO-FFNN is to improve the training and performance of the FFNN by using the GWO algorithm to optimize the neural network's weights and biases. The SGWO-FFNN has been shown to be effective in various applications, such as image classification, time series prediction, and function approximation.

Keywords: *Crop Yield, Prediction, Classification, Grey Wolf, Neural Network, Optimization*

1. INTRODUCTION

The capacity of a country to feed itself is a crucial factor in determining its success. The cultivation of vital food crops has always been linked to farmers. Our culture's primary focus is the high population expansion rate[1]. There has been a significant reduction in agricultural potential, especially regarding land productivity and use. Because the amount of arable land is unlikely to expand in this age of urbanization and globalization, more attention must be paid to optimizing the already available land. Predicting which crop cultivars will be successful is essential in the agricultural industry. While new studies have increased access to agricultural statistics, few have considered the feasibility of making accurate crop predictions using past data[2]. However, it is difficult to estimate crop cultivation due to the unrestrained use of fertilizers, including nitrogen, potassium, and micronutrients. Soil texture, precipitation, and temperature are all agro-climatic input characteristics that impact crop yield. Collecting such data across vast areas is challenging since agricultural input factors differ from region to region. The large collected datasets have the potential for widespread use in crop prediction. The complexity of the issues necessitates the creation of novel machine-learning approaches for cultivating arable land and optimizing limited land resources. Agriculturalists have been experimenting with various forecasting methods to determine which crop suits a given plot of land[3].

Agriculture is deeply intertwined with the history and identity of a nation. Farmers worldwide confront various difficulties due to various factors, including but not limited to: storms, floods, land drifting, soil erosion, poor soil quality, and erratic weather[4]. Due to the country's enormous population, agriculture provides the foundation for India's entire manufacturing sector. Agricultural failure has far-reaching effects on the productivity of other sectors, lowering the country's gross domestic product[5]. For this reason, bolstering agriculture is crucial to the country's economic development, industrial balance, and price stability. Farmers in developing nations lack the technical knowledge and education to keep up with the rapid pace of change in the developed world. Farmers don't know about new farming strategies,

equipment, and techniques available[6]. They aren't aware of genetically modified seeds, which increase harvest yields significantly. More preventive actions are necessary for farmers to increase their technical skills to keep up with the rapid pace of technological change. Choosing a crop well-suited to the farmland is crucial to the prosperity of the agricultural enterprise. It is impossible to make a sound crop selection choice based on a single criterion alone; instead, several considerations must be considered[7]. When making a choice that must consider more than one factor, a multi-criteria decision-making (MCDM) technique is used. In the context of crop selection choices in agriculture, several variables are weighted more heavily than others[8].

Regional and worldwide crop production forecasting is essential for crop producers, agricultural managers, food security alerts, food trade policies, and carbon cycle studies[9]. To provide a steady food supply for its citizens, governments in nations with a high population density and a scarcity of farmland must prioritize the development of dynamic and accurate crop forecasting systems. To maximize profits for local farmers and the national export sector, countries with a small population but a lot of farmland are more focused on the effects of actual crop output on food commerce[10].

Bio-inspired optimization is a method of solving problems inspired by natural processes and mechanisms[11]. This approach can be applied to various fields, including machine learning (ML) and pattern recognition. In the field of classification, bio-inspired optimization can be used to improve the performance of classifiers by optimizing the parameters of the classifier using techniques such as genetic algorithms, particle swarm optimization, and ant colony optimization[12]. These bio-inspired algorithms can be used to find the optimal set of parameters for a given classifier, resulting in improved accuracy and robustness in all different domains, including advanced networking [13]–[24], [25]. Some of the advantages of Bio-Inspired Optimization Algorithms (BIOA) are listed below:

1. BIOA is based on natural processes and systems, which can provide better solutions to complex problems than traditional methods.
2. These algorithms can handle high-dimensional and non-linear data, which is common in crop yield prediction.
3. They are less sensitive to initial conditions and parameters, which makes them more robust than traditional methods.
4. BIOA are parallelizable, meaning they can use modern computing resources to speed up the computation process.
5. They can also be easily integrated with other machine-learning techniques to improve the performance of crop yield prediction models.
6. They are highly flexible, allowing for the incorporation of new data and knowledge as it becomes available.
7. They can handle missing or incomplete data, which is common in crop yield prediction.
8. They can also be used to optimize multiple objectives and constraints, which is vital in crop yield prediction as it often involves balancing factors such as yield, water usage, and pest resistance.

1.1 Problem Statement

Several issues can arise when using ML to predict crop yields, including:

1. Data availability and quality: A large amount of data is required to train an ML model for crop yield prediction. However, obtaining high-quality data can be difficult, particularly in developing countries where data collection and infrastructure may be limited.
2. Feature selection: Crop yield prediction models often rely on many input features, such as weather data, soil data, and historical crop yields. Selecting the most informative features can be challenging and may require domain-specific knowledge.
3. Model complexity: Crop yield prediction models with many interacting variables can be complex. This can make understanding how a model makes its predictions difficult and increase the risk of overfitting.
4. Spatial and temporal variability: Crop yields can vary widely depending on location and time of year. It isn't easy to develop models that generalize well to different regions and seasons
5. Unforeseen events: Climate change, pandemics, natural disasters, and other unforeseen events can make crop yield predictions more difficult.
6. Model interpretability: In some cases, ML models may be too complex to interpret, making it difficult to understand why a model is making a particular prediction.
7. Human factors: Farmer's practices, technological advancements, and government policies can also impact crop yields, making it difficult to predict with machine learning.

8. Soil variability: Variations in soil quality and fertility can significantly impact crop yields.
9. Weather and climate variability: Weather and climate can significantly impact crop yields and can be difficult to predict. Incorporating weather and climate data into the model can be challenging.
10. Data availability: Deep learning models require large amounts of data to train effectively, and data on crop yields can be challenging to obtain in specific regions or for certain crops.
11. High computational requirements: Training ML models can require significant computational resources, making it difficult for researchers to access the necessary hardware.
12. Limited understanding of crop physiology: The physiological processes that drive crop growth are not well understood, making it challenging to develop accurate predictions of yields.

1.2 Motivation

Following is the list of things that motivated this research work:

1. Improving crop productivity and food security
2. Identifying and mitigating the effects of climate change on crop yields
3. Developing more efficient and sustainable farming practices
4. Predicting and managing crop disease and pest outbreaks
5. Enhancing the accuracy and reliability of crop yield predictions
6. Supporting farmers and agribusinesses in decision-making and resource allocation.
7. Understanding the relationship between crop yield and environmental factors
8. Developing precision agriculture techniques.
9. Optimizing crop growth and yield under different weather conditions.
10. Identifying the genetic and physiological factors that contribute to crop yield.

1.3 Research Objective

The research objective of "Strenuous Grey Wolf Optimization-based Feed-Forward Neural Network (SGWO-FFNN) for Enhanced Crop Yield Prediction" is to develop and evaluate a novel optimization algorithm, the Strenuous Grey Wolf Optimization (SGWO) algorithm, for training feed-forward neural networks (FFNNs) which will result in enhanced crop yield prediction, and to demonstrate the effectiveness of the proposed approach through extensive experiments and comparison with state-of-the-art methods.

1.4 Organization of the Paper

Section 1 summarized the Introduction with a clear and concise statement of the problem being addressed in the research paper, an explanation of why the problem is important and relevant, and a statement of the specific objectives of the research. Section 2 critically reviewsthe existing literature on the topic, including relevant research studies, theories, and models. Section 3 provides a detailed description of the proposed classifier, including its architecture, algorithms, and any novel contributions. Section 4 illustrates the description of the dataset used for the research, including its size and features. Section 5 describes the metrics used to evaluate the performance of the classifiers. Section 6 presents a discussion of the results obtained from the classifiers, including any comparison with other existing methods. Section 7 summarizes the main findings and contributions of the research, as well as any limitations and future work.

2. LITERATURE REVIEW

A literature review on crop yield prediction would involve researching and summarizing existing studies and papers on the topic. This could include examining different methods and models for predicting crop yields, analyzing their accuracy and limitations, and identifying gaps in current knowledge and areas for future research. Some key focus areas could include using meteorological data, remote sensing, and ML techniques to predict crop yields and the impact of changes in climate and other environmental factors on crop yields. The literature review might also examine the practical applications of crop yield prediction, such as in agriculture, food security, and biofuels.

"DeepYield"[26] has been suggested for predicting crop production. It combines different types of neural networks, specifically Convolutional Long Short-Term Memory layers and a 3D CNN, to improve the extraction of features related to space and time. The models are trained using various types of data, including information on land cover, surface reflectance, temperature, and historical crop yields from 1836 countries in the US where soybeans are primarily grown.A method called "Synthetic Analysis and Bootstrap Method (SABM)"[27] has been put forth for forecasting worldwide crop yields. It employs updated information from crop surveys for around 12,000 geographic regions. To begin, it examines the concurrent effect of El Nino Southern

Oscillation (ENSO) on the yield fluctuation of key crops using SABM. Next, it calculates the likelihood of simultaneous crop failure in the top five nations that produce the most crops by utilizing a statistical technique called the copula approach. Lastly, using multiple regression analysis, the most accurate forecasting model, relevant ENSO indices, and the appropriate lead time for each geographical area are determined based on previous ENSO indices."Bayesian-based Model" [28] is proposed to forecast the results of wheat breeding experiments. This model utilizes hyperspectral reflectance data from the canopy of the plants and employs the ensemble Bayesian model averaging technique to enhance its accuracy. To construct yield prediction models, vital spectral bands identified using the Boruta feature selection technique are input into a combination of four linear and four non-linear ML models. These predictions are then combined using Bayesian model averaging weights, which are determined based on the cross-validation performance of each model.

"Soft-Fruit Yield Framework"[5] has been developed to predict soft fruit yield. It incorporates weather forecasts at a regional scale and utilizes autonomous sensors to gather data on local crop conditions, specifically focused on strawberries as a model crop. The study involved planting seedlings in polytunnels and collecting data on various environmental factors and yield throughout the growing season. Over 1.2 million data points were gathered using microsensors that measured the temperature, humidity, soil moisture, and light levels within the polytunnels. To enhance the study, some plants were covered with fleeces to create more variation within the polytunnels. The data was then analyzed using trigonometric models, which helped to improve the accuracy of temperature and humidity predictions specific to the polytunnels by adjusting for discrepancies between the weather station data and the data gathered by the microsensors." Linear Regression-based Classification"[9] has been suggested for forecasting the characteristics of soil and the output of corn crops. This method involves the creation of multiple models through linear regression and assessing five different ML techniques based on their coefficient of determination and root mean square error. Combining remotely sensed data with ML techniques is a viable approach for identifying soil properties and corn yield at a regional level, which can aid in identifying regions that may require special attention and implementing farming methods specific to those areas. "Bayesian Extreme Gradient Boosting"[29] has been suggested to determine the impact of factors such as crop management, soil, and weather on the predicted yield and the level of uncertainty for key components of maize production. This approach combines Bayesian statistics to model the yield response and extreme gradient boosting to evaluate the importance of different features in predicting the outcome. The study found that factors related to crop management, such as the previous crop and irrigation, have a significant impact (around 50%) on estimating one component of the yield but a lesser impact on other components.

Crop production data from 40 agricultural areas were organized into fewer, larger regions for forecasting using the "Hierarchical Clustering-based Prediction Model (HCPMM)" [30]. Model-based recursive partitioning, linear regression, and Bayesian neural networks predicted agricultural yields. The mean absolute error skill score from 2000 to 2011 was used to assess predictions generated using data from July and earlier. A "Comparison Study" [31] used two significant ML strategies, namely, Random Forest and Support Vector Machine, to a commonly used model for predicting the conductance of stomata in wheat plants. The study used data collected from an experimental facility in Australia. The results showed that the ML methods were more accurate than the traditional model in predicting conductance based on various environmental factors. However, the ML methods required a large amount of data for training and did not offer the same level of understanding of plant physiology as the traditional model."Hybrid Particle Swarm Optimization with Extreme Learning Machine" [32] has been proposed for forecasting evapotranspiration in regions with limited climate information. This method combines a particle swarm optimization algorithm and an extreme learning machine model to optimize the model's parameters. The method was tested in northwest China's arid region and used three input data sets: radiation, temperature, and mass transfer."Multilayer Perceptron Model with Bayesian and Copula Bayesian Approaches" [33] for predicting wheat yields was proposed that utilizes a Multilayer Perceptron (MLP) model in multiple ways. The model was trained using backpropagation in its default mode and with a combination of the Genetic Algorithm, Sine-Cosine Algorithm, and Water Striders Algorithm in a hybrid mode. The ensemble modeling techniques employed are Copula-based Bayesian Model Averaging and Bayesian Model Averaging.

"Random Forest Algorithm (RFA)" [34] is a popular method for tackling both classification and regression tasks in supervised machine learning. It operates by building a collection of decision trees, known as a forest, and then using the average or majority vote of the predictions made by each tree to make a final prediction. This approach helps to reduce the variance and bias that can arise when using just a single decision tree. Additionally, RFA employs random selection of features at each split, which helps to avoid overfitting. This algorithm is versatile in that it can handle both categorical and numerical features, and it can also identify the most important features in a dataset. RFA is commonly used in a variety of applications, such as image classification, speech recognition, and bioinformatics. One of its main benefits is its ability to handle large amounts of data and high-dimensional feature spaces. However, it should be noted that RFA can be computationally expensive and may not be the best option for real-time applications. In summary, RFA is a robust algorithm that can produce accurate predictions and feature importance, but its performance may not be optimal when used to predict crop yields. "Support Vector Machine (SVM)" [35] is a powerful supervised learning algorithm that is widely used for both classification and regression tasks. The algorithm is designed to find the optimal boundary or hyperplane that separates the data into different classes, while maximizing the margin, which is the distance between the boundary and the closest data points from each class, known as support vectors. One of the key advantages of SVM is its ability to handle non-linearly separable data through a technique called kernel trick, which projects the data into a higher dimensional space where it becomes linearly separable. Furthermore, SVM is robust to high-dimensional data and can handle a large number of features. However, SVM has some limitations as well, such as being sensitive to the choice of kernel function and algorithm parameters, and being computationally expensive for large datasets. In the context of crop yield prediction, it has been observed that the use of SVM may lead to a decline in performance.

3. STRENUOUS GREY WOLF OPTIMIZATION-BASED FEED-FORWARD NEURAL NETWORK (SGWO-FFNN)

3.1. Feed Forward Neutral Network

Bernstein polynomials uniformly approximate a limited and continuous function over the range [0, 1]. It is defined as follows: the polynomial function $g(p)$ for which $p\ \omega\ [0,1]$ and $\vartheta\ \omega\ T^*$ in terms of a Bernstein polynomial of order ϑ of $g(p)$, with a binomial coefficient denoted by $U_\vartheta^\ni = \ni !/(\vartheta!\,(\vartheta-\ni)!)$. $g(p)$ on [0,1] is mapped to a polynomial function $V_\vartheta(g,p)$ with $p\ \omega$ [0, 1] through the polynomial function, and it is expressed as Eq.(1).

$$V_\vartheta(g,p) = \sum_{\ni=0}^{\vartheta} g\left(\frac{\ni}{\vartheta}\right) U_\vartheta^\ni p^\ni\,(1-p)^{\vartheta-\ni} \qquad (1)$$

There exist T parameters in $g(.)$, this research work writes it as $g(P) = [p_1 p_2 \ldots . p_T]^F\ \omega\ [0,1]^{T\times 1}$, where the F in the superscript denotes the transposition operator. Therefore, Eq.(1) can be written as Eq.(2).

$$V_{\vartheta,T}(g,P) = \sum_{\ni_1=0}^{\vartheta} \ldots \sum_{\ni_T=0}^{\vartheta} g\left(\frac{\ni_1}{\vartheta}, \ldots ., \frac{\ni_T}{\vartheta}\right) v_{\ni_1}(p_1) \ldots . v_{\ni_T}(p_T), \qquad (2)$$

such that $v_{\ni_s}(p_s) = U_\vartheta^{\ni_s} p_s^{\ni_s}(1-p_s)^{\vartheta-\ni_s}$ with $s = 1,2,\ldots,T$.

Regarding Approximation Theory [1], Chebyshev Polynomials ($ChsPlm$) [2] are a sequence of basis functions that prove useful. ANNs trained with $ChsPlm$ perform well in approximation and generalization. Eq.(3) determine the recursion.

$$\begin{aligned} F_0(p) &= 1, \\ F_1(p) &= p, \\ F_{w+1}(p) &= 2pF_w(p) - F_{w-1}(p), \end{aligned} \qquad (3)$$

wherein $ChsPlm$ severity is indicated by $w = 1,2,3\ldots,n$.

An assortment of $ChsPlm$s, in conformity with the theory of orthonormal approximation, can estimate a target function of arbitrarily high degree $m(p)$ with $p\omega$ in the range $[-1,+1]$ as Eq.(4).

$$m(p) \approx \sum_{z=0}^{Z} \delta_z F_z(p), \qquad (4)$$

wherein Z is the sum of the $ChsPlm$s employed in approximating $m(p)$, and δ_z is the mass of $F_z(p)$. To simulate each $v_{\ni_s}(p_s)$ in Eq.(2), this research work utilizes a cluster of $ChsPlm$s when Z is sufficiently a huge value and it is represented as Eq.(5)

$$v_{\ni_s}(p_s) \approx \sum_{z_s=0}^{Z_{s,\ni_s}} \delta z_{s,s,\ni_s} F_{z_s}(p), \quad (5)$$

using $s = 1,2,\ldots,T$, and if we let $M_{\ni_{1,\ldots\ldots,}\ni_T}(P) = v_{\ni_1}(p_1)\ldots\ldots v_{\ni_T}(p_T)$, we get:

$$\begin{aligned} & P_{\ni_{1,\ldots,}\ni_T}(P) \\ & \approx \left(\sum_{z_1=0}^{Z_{1,\ni_1}} \delta z_{1,1,\ni_1} F_{\ni_1}(p_1)\right) \ldots \left(\sum_{z_T=0}^{Z_{T,\ni_T}} \delta z_{T,T,\ni_T} F_{z_T}(p_T)\right) \\ & = \sum_{z_1=0}^{Z_{1,\ni_1}} \ldots \sum_{z_T=0}^{Z_{T,\ni_T}} \delta_{z_1,1,\ni_1} \ldots \delta z_{T,T,\ni_T} F_{z_1}(p_1) \ldots F_{z_T}(p_T). \end{aligned} \quad (6)$$

Since, $ChsPlm$s are approximative, Eq.(2) can be represented as Eq.(7):

$$\begin{aligned} & V_{\vartheta,T}(g,P) \\ & = \sum_{\ni_1=0}^{\vartheta} \ldots \sum_{\ni_T=0}^{\vartheta} g\left(\frac{\ni_1}{\vartheta}, \ldots\ldots, \frac{\ni_T}{\vartheta}\right) M_{\vartheta_{1,\ldots\ldots}\ni_T}(P) \\ & \approx \sum_{z_1=0}^{z_1^{(max)}} \ldots \sum_{z_T=0}^{z_1^{(max)}} n_{z_{s,\ldots,}z_T} F_{z_1}(p_1) \ldots F_{z_T}(p_T) \\ & = \sum_{a=1}^{A} n_a G_a(P), \end{aligned} \quad (7)$$

wherein $\{g_A(p|A = 1,2,\ldots,A)\}$ signifies the basis function, $\{F_{z_1}(p_1)\ldots\ldots F_{z_T}(p_T)\}, n_a$ denotes the weight associated with each $G_a(P)$, & A is the maximum number of basis functions. Each $G_a(P)$is increased by some number of $ChsPlm$s equal to T. These $ChsPlm$s can be sorted according to their degrees to help clear things up. The lexicographic order is chosen in this research work by upgrading it to categorize $ChsPlm$s.

3.1.1 Model of the original FFNN's

It is possible to represent a real-world instance with T characteristics as $P_{ins} = [p_1 p_2 \ldots . p_T]^F \omega\ [0,1]^{T\times 1}$. Then, by using the FFNN model, $Q_{ins} = [q_1 q_2 \ldots\ldots q_C]^F \omega\ B^{C\times 1}$can be derived, where C stands for the total count of classes. If the class $d(1 \le w \le C)$, then $q_d = 2$ and $q_w = 1(1 \le w \le C\ and\ w \ne d)$. This is the result of each $q_s(s = 1,2,\ldots,C)$in Q_{ins}.

Remembering that the base FFNN model incorporates C subnets, each of which may be thought of as a standard three-layer neural network. The neural network can be represented for the uth subnet, where $u = 1,2,\ldots,C$. To represent the T characteristics with the output result y, the subnet model depicted holding T neurons in its input layer and a single neuron in its output layer.

The hidden layer neurons of the modified FFNN are powered up by a series of decided-to-order basis functions $\{G_{u,a}(P)|a = 1,2,\ldots,A\}$, and the amount of these neurons A is deduced using a significant classifier. The linking weights between the input and hidden layers are maintained at 1 to reduce computational complexity without sacrificing classification accuracy. With important parameters, the weights that link the hidden layer to the output layer $\{n_{u,a}(P)|a = 1,2,\ldots,A\}$, are calculated using the weights-based classification strategy (WCS). In addition, every neuron has a constant threshold value of 0. Eq.(8) depicts the expected outcome of the subnet.

$$q_u(P) = \sum_{a=1}^{A} n_{u,a} G_{u,a}(P). \quad (8)$$

Approximation and generalization performances, crucial indicators of a neural network's quality, are strongly influenced by the count of neurons present in it. It's necessary to compute the neurons present in the hidden layer because the percentage of neurons in each layer is equal to the number of characteristics of the true instance, and the amount of neurons in the output layer is equal to the number of records present in the dataset.

The mean square error (MSE) is frequently employed to evaluate the efficacy of neural networks. It's a measure of how much the neural network's actual outputs deviate from what would be predicted given the inputs to the network. MSE is determined using Eq.(9) when the training dataset comprises X instances that may be treated as $\{Q_s|s = 1,2,\ldots,X\}$.

$$MSE_u = \sqrt{\frac{1}{X}\sum_{s=1}^{X}\left(\hat{q}_{u,s} - q_u(P_s)\right)^2}, \quad (9)$$

wherein $\hat{q}_{u,s}$ indicates the sth desired output (i.e., $\hat{q}_{u,s} = 2$ if the sth example belongs to class u; otherwise, $\hat{q}_{u,s} = 1$), and $q_u(P_s)$ is the real weight of the corresponding sth instance's output attribute. It's worth noting that a lower MSE suggests a closer approximation between the neural network's prediction and the desired prediction.

To get enhanced classification accuracy, partitioning becomes mandatory and the training dataset into n different sets is then used in n iterations of cross-validation. With $y = 1,2,\ldots,n$, the y^{th} iteration applies the y^{th} subset to validation while the remaining subsets are utilized for training. The algorithm is executed with a single hidden-layer neuron, and that number will rise by

one each time. Furthermore, the MSE of the verification subset is logged as H_{val}as well as the MSE of the training section is recorded as H_{tra} while the algorithm is being executed.

The early-stage approximation and generalization ability of the neural network improves as the number of hidden-layer neurons, as seen by the constant lowering of H_{val} and H_{tra}. Overfitting occurs when a neural network's hidden layer has multiple neurons, leading to subpar approximation and generalization performance. This means that as the total number of hidden layer neurons increases to a significant size, the ratio of H_{tra} to H_{val} keeps falling while it attempts to get increase. It can be stated that once the H_{val}starts to increase continuously over the threshold value of neurons in the hidden layer. The number of neurons correlating to the minima of H_{val} is presumed to be the optimal amount in the cross-validation stage.

3.1.2 Weight-based Classification Strategy

Once the number of hidden layer neurons is set, then the topology of the neural network is fixed by depending on it. It involves calculating the weights of the connections between both the hidden and the output layer $\{n_{u,a}(P)|a = 1,2, \dots., A\}$. The weights of the uth FFNN subnet are represented in Eq.(10) as a vector.

$$N_u = \left[n_{u,1} n_{u,2} \dots.. n_{u,A}\right]^F \omega\, B^{A\times 1}. \quad (10)$$

Eq.(11) defines the input activation matrix.

$$S_u = \begin{bmatrix} G_{u,1}(P_1)G_{u,2}(P_1) & \dots & G_{u,A}(P_1) \\ G_{u,1}(P_2)G_{u,2}(P_2) & \dots & G_{u,A}(P_2) \\ \vdots & \ddots & \vdots \\ G_{u,1}(P_X)G_{u,2}(P_X) & \dots & G_{u,A}(P_X) \end{bmatrix} \omega\, B^{X\times A}. \quad (11)$$

The desired $\{\hat{q}_{u,s}|s = 1,2, \dots., X\}$from the neural net is expressed as Eq.(12) as a vector.

$$Q_u = \left[\hat{q}_{u,1}\hat{q}_{u,2} \dots \hat{q}_{u,X}\right]^F \omega\, B^{X\times 1}. \quad (12)$$

Then, the WCS assists in deriving the optimum values using Eq.(13) of the subdomain of the modified FFNN.

$$N_u = (S_u^F S_u)^{-1} S_u^F Q_u = pinv(S_u) Q_u, \quad (13)$$

Using the notation $pinv(S_u)$to represent the effects of the matrices S_u. Using the same technique, parameters for the other FFNN subnets can be obtained but may result in low classification accuracy then it is expected. Hence, this research work utilizes the enhanced version of grey wolf optimization, namely "Strenuous Grey Wolf Optimization" to enhance the classification accuracy.

3.2Strenuous Grey Wolf Optimization

This part introduces the grey wolf optimizer's origin, the underlying mathematical model, and the associated method. There is a fairly clear pecking order among grey wolves. The most common digit sequences are δ, γ, θ and π. A wolf with level δ is the team's leader. Subordinate γ wolves of the 2nd layer help the alpha wolf with decision-making. Wolves in the roles of Scout, guardians, seniors, hunters, and caretakers make up the γ who are subservient to the δ, γ who is in charge of the π. The pack's lowest-ranking wolf is π. They are in the last rank in the pack. Among some communities, the π wolf serves as a babysitter. It's important to note that while developing a computational formula for hunting grey wolves, pick the three optimum places to be alpha (δ), beta (γ), and delta(θ). We narrowed it down to π as the last possible choice, and alpha(δ), beta(γ), and delta(θ) helped steer our hunting position. This is not to say that there is just one alpha (δ) wolf, or even one beta (γ)wolf, or even one delta(θ) wolf.

Grey wolves follow a three-stage process when hunting: surrounding prey, hunting, and attacking prey (exploit). Each wolf adjusts its position relative to the herd, dependent on how close it is to the prey. A siege is eventually organized. Eq.(14) and Eq.(15) mathematically show how the wolves encircle the prey.

$$\vec{Y} = \left|\vec{UD}.\overrightarrow{P_m}(f) - \vec{P}_f\right| \quad (14)$$

$$\vec{P}_{(f+1)} = \overrightarrow{P_m}(f) - \vec{D}.\vec{Y} \quad (15)$$

$\vec{Y}$ shows how far away the prey is from the wolf. $\overrightarrow{P_m}(f)$ shows the prey's position vector. $\overrightarrow{P_m}$ is the vector that shows where each wolf is.

Eq.(16) and Eq.(17) determine the correlation coefficient vectors $\vec{D}$ and $\vec{U}$.

$$\vec{D} = 2\vec{d}.\vec{b}_1 - \vec{d} \quad (16)$$

$$\vec{U} = 2.\vec{b}_2 \quad (17)$$

The random vectors $\vec{b}_1$ and $\vec{b}_2$ are between [0, 1]. From 2 to 0, $\vec{d}$ reduces linearly as the amount of iterations increases.

3.2.1 Hunting

alpha(δ), beta(γ), and delta(θ) are more likely to know where to find suitable prey. Following these three wolves, the others adjust their locations accordingly. Eq.(18) to Eq.(20) indicates the mathematical model of the same.

$$\vec{Y}_d = \left|\vec{U}_1\hat{D}.\vec{P}_\delta - \vec{P}\right|\vec{Y}_\gamma = \left|\vec{U}_2\hat{D}.\vec{P}_\gamma - \vec{P}\right|\vec{Y}_\theta = \left|\vec{U}_3\hat{D}.\vec{P}_\theta - \vec{P}\right| \quad (18)$$

$$\vec{P}_1 = \vec{P}_\delta - \vec{D}_1.(\vec{Y}_\delta)\vec{P}_2 = \vec{P}_\gamma - \vec{D}_2.(\vec{Y}_\gamma)\vec{P}_3 = \vec{P}_\theta - \vec{D}_3.(\vec{Y}_\theta) \quad (19)$$

$$\vec{P}(f+1) = \frac{(\vec{P}_1 + \vec{P}_2 + \vec{P}_3)}{3} \quad (20)$$

If $\vec{Y}_\delta, \vec{Y}_\gamma, \vec{Y}_\theta$ denote the search agent's position, $\delta, \gamma, \theta, \vec{P}_\delta, \vec{P}_\gamma, \vec{P}_\theta$ denote the wolves' locations, and δ, γ and θ denote the search agent's distance from the wolves. So far, those are the three most promising options. Wolf packs are given a random weight by $\delta, \gamma, \theta,$ the correlation coefficient, so they do not rush at their prey and reach sexual maturity at an unnaturally young age. It plays the same part that barriers do in a real hunt. After $(f+1)$ iterations, the search agent is located at the position $\vec{P}(f+1)$. You can see the three wolves at $\delta, \gamma,$ and θ and have a significant impact on its position update.

Algorithm 1. Strenuous Grey Wolf Optimizer

1	Grey wolf population $P_s(s = 1,2,3, \dots, t)$ initialized.
2	Prepare D, d, U
3	Find out how effective each search agent is.
4	Describe δ, γ, θ, wolves
5	Primary loop
6	Analyze $(f < F)$
7	For each search agent
8	Adjust the locations using Eq (20)
9	End foreach
10	Upgrade D, d, U
11	Calculate each search agent's fitness.
12	Upgradeδ, γ, θ
13	$f = f + 1$
14	Finish while
15	Repeat going to Step 5 till a better result is achieved.

3.2.2 Exploring

Once their target comes to a stop, the wolves pounce. The process can be modulated by reducing $\vec{d}$. From Eq.(3), this research work can deduce that the range of values for $\vec{D}_1, \vec{D}_2, \vec{D}_3$ is $[-2,2]$, and that as $\vec{d}$, their variability progressively diminishes. Grey wolves prioritize exploitation when $|D| < 1$. During periods of $|D| > 1$, when grey wolves avoid their usual prey in favor of searching for something new, foraging takes centre stage.

4. ABOUT THE DATASET

A dataset in machine learning research is a collection of data that is used to train, validate, and test a model.

4.1 Soybean Crop Dataset

The dataset used for this work contains information regarding the Soybean crop cultivated along with the nine states of the US Corn belt. The dataset holds 25345 records and three different features, namely: (a) Soil Feature, (b) Weather Feature, and (c) Crop Management Feature.

4.1.1 Soil Feature

The data present in this feature contains 11 variables measured at six different depths in 250 square meter resolution, which are: (i) 0-5cm, (ii) 5-15cm, (iii) 15-30cm, (iv) 30-60cm, (v) 60-100cm, and (vi) 100-200cm.

Table 1: Six Different Depths

Interval	Top Depth(cm)	Bottom Depth(cm)
I	0	5
II	5	15
III	15	30
IV	30	60
V	60	100
VI	100	200

Table 2:Soil Features

Features	Units	Description
Bulk Density (bdod)	Centigram Per Cubic Centimetre	Density in a volume of the fine-grained earth
Cation Exchange Capacity (cec)	Millimole Per Kilogram	Soil pH7 CEC: Cation Exchange Capacity (i.e., the potential

		of Hydrogen value-7)
Coarse Fragments (cfvo)	Cubic Centimetre Per Cubic Decimetre	The proportion of coarse particles (> 2 mm) by volume.
Clay	Gram Per Kilogram	Percentage of clay particles in the fine earth fraction
Total Nitrogen (Nitrogen)	Centigram Per Kilogram	N_2O (i.e., the Entire Gas)
Organic Carbon Density (ocd)	Kilogram Per Cubic Meter	The molecular weight of carbon in organic compounds
Organic Carbon Stock (ocs)	Kilogram Per Square Meter	Carbon Remaining in Organic Matter
pH in H_20	Percentage	Hydrogen's Untapped Potential in Water
Sand	Gram Per Kilogram	Sand's representation in the fine earth's particle distribution
Silt	Gram Per Kilogram	Microscopic particles of silt relative to those of fine earth
Soil Organic Carbon (soc)	Decigram Per Kilogram	The fine-grained organic carbon content of the soil

4.1.2 Weather Feature

The data present in this feature contains five components which are measured throughout the year.

Table 3:Weather Features

Features	Units	Description
Precipitation	Millimetre	Week typical precipitation
Solar Radiation	Watt Per Square Meter	Surfaces are exposed to solar energy
Maximum Temperature	Degree Centigrade	Weekly mean highest temperature
Minimum Temperature	Degree Centigrade	Weekly mean low temperature
Vapor Pressure	Pascal	Speed of a liquid's evaporation

4.1.3 Crop Management Feature

Information on the cumulative proportion of planted fields each state has each week beginning in April of each year is provided in this feature.

Table 4:Crop Management Features

Features	Units	Description
Weekly Cumulative Percentage of Planted-fields (CWPP)	Percentage	The average weekly proportion of fields that have been planted

4.2 Rice Crop Dataset

This study incorporated a dataset that pertains to rice farming in 20 districts of Tamilnadu, spanning from 1990 to 2015. It comprises 520 records and encompasses two distinct characteristics: soil and weather. To gain a comprehensive understanding of classifier performance, records in the dataset were replicated multiple times, resulting in 10400 records.

4.2.1 Soil Feature

The data set for soil examination includes measurements for six attributes, taken monthly for 25 years starting in 1990.

Table 5:Soil Feature

Features	Units	Description
Irrigated area of rice	Ha	Rice Irrigated Area
Consumption of Nitrogen	Tons	NitrogenFertilizer
Consumption of Potassium	Tons	PhosphateFertilizer
Consumption of Potash	Tons	PotashFertilizer
Presence of OrthidSoil	Percentage	Orthid Soil Level

Presence of Sandy Soil	Percentage	Sandy Alfisol Soil Level

4.2.2 Weather Feature

The measurement of the 5 components of the weather were recorded during a 12-month period between 1990 and 2015.

Table 6:Weather Feature

Features	Units	Description
JAN_MR - DEC_MR	mm	Rainfall in Months
JAN_MT - DEC_MT	cg	Temperature at Minimum Level
JAN_XT - DEC_XT	cg	Temperature at Maximum Level
JAN_PT - DEC_PT	mm	Precipitation
JAN_ET - DEC_ET	mm	Evapotranspiration Potential

5. PERFORMANCE METRICS

The metric outlined below is employed in this research to determine the effectiveness of the proposed classifier when compared to the current classifier.

- **Classification Accuracy (CA)** measures how well a classifier can correctly identify the class of a given input. It is calculated as the number of correct classifications divided by the total number of instances in the test set.
- **F-Measure(FM)**also known as the F1 Score, is a metric used to evaluate the performance of binary classification models. It is calculated by taking the harmonic mean of precision and recall. Precision is determined by dividing the number of true positive results by the number of true positive and false positive results. Recall is found by dividing the number of true positive results by the number of true positive and false negative results.
- **Matthews Correlation Coefficient (MCC)** is another metric for evaluating binary classification models. It measures the correlation between predicted and observed outcomes, taking into account true positives, true negatives, false positives, and false negatives. A value of 1 indicates a perfect prediction, while a value of 0 suggests the model is no better than random guessing. A value of -1 indicates complete disagreement between the predictions and observations.
- **Fowlkes-Mallows Index (FMI)** is a similarity index that is used to compare the similarity of two clusterings. It measures the geometric mean of the pairwise precision and recall of the two clusterings, where precision is the number of true positives divided by the number of true positives and false positives, and recall is the number of true positives divided by the number of true positives and false negatives.

6. RESULTS AND DISCUSSION

6.1 CA and FM Analysis

Figures 1 and Figure 2 present an analysis of how different classifiers perform when using soybean and rice crop datasets, measured in terms of CA and FM. The figures indicate that the proposed classifier outperforms the others because of its optimization strategy, while the RFA and SVM classifiers fall short because they do not have an optimization strategy. These results are also listed in Table 7. Thus the proposed work achieves the research objective of predicting the crop yields more accurately than the existing classifier.

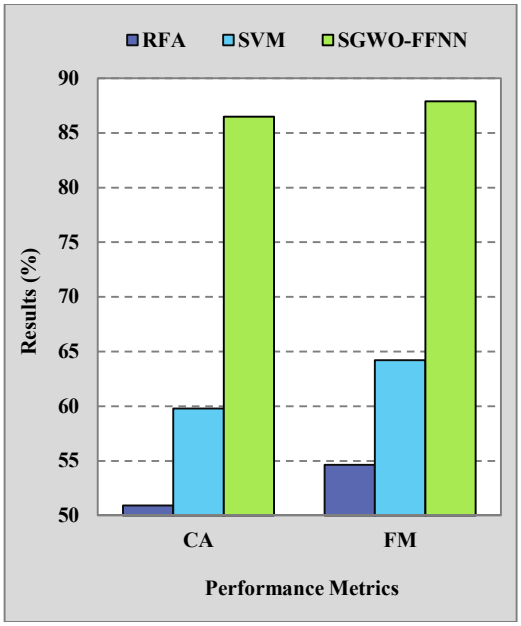

Figure 1. Analysis of CA and FM on Soybean Crop Dataset

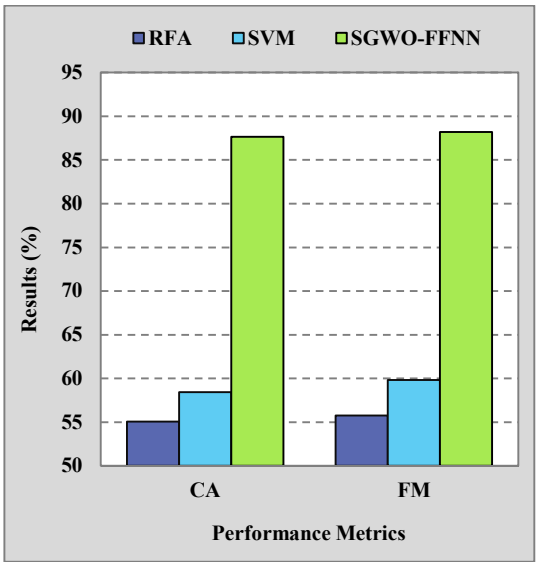

Figure 2. Analysis of CA and FM on Rice Crop Dataset

Table 7 Result Values of CA and FM

Dataset →	Soybean Crop Dataset		Rice Crop Dataset	
Metrics→ Classifiers↓	CA	FM	CA	FM
RFA	50.925	54.632	55.077	55.749
SVM	59.795	64.210	58.442	59.818
SGWO-FFNN	86.475	87.884	87.654	88.179

6.2 FMI and MCC Analysis

Figure 3 and Figure 4 provide an analysis of the performance of classifiers using soybean and rice crop datasets with respect to FMI and MCC. The results of the study suggest that the proposed classifier is trustworthy and can be utilized in further processing. In contrast, the results obtained from RFA and SVM classification methods indicate that further developments are required. The data values used in Figure 3 and Figure 4 are detailed in Table 8. FMI and MCC Analysis indicate that the proposed classifier's result is more trustable than the existing classifiers.

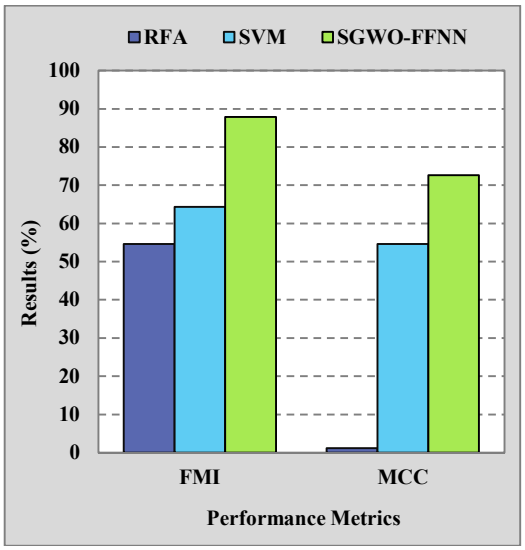

Figure 3. Analysis of FMI and MCC on Soybean Crop Dataset

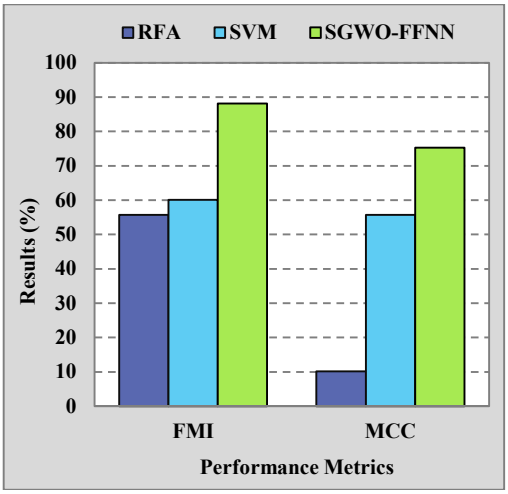

Figure 4. Analysis of FMI and MCC on Rice Crop Dataset

Table 8 Result Values of CA and FM

Dataset →	Soybean Crop Dataset		Rice Crop Dataset	
Metrics→ Classifiers↓	FMI	MCC	FMI	MCC
RFA	54.633	1.194	55.750	10.135
SVM	64.336	54.633	60.060	55.750
SGWO-FFNN	87.885	72.587	88.179	75.259

7. CONCLUSION

Crop yield prediction is a critical task for farmers and agricultural scientists alike. In recent years, ML algorithms have succeeded in this

domain. The ML algorithm can be used to predict crop yields. These algorithms learn from past data to create a model that can then be used to predict future crop yields. Many different classification algorithms are available, each with its strengths and weaknesses. The Strenuous Grey Wolf Optimization-based Feed-forward Neural Network (SGWO-FFNN) has been successfully applied in crop yield prediction. The results of this study indicate that the proposed model can effectively predict crop yields with a high degree of accuracy. The use of the SGWO algorithm in optimizing the FFNN model has been shown to improve the network's performance. The SGWO demonstrate the potential of the SGWO-FFNN model for accurate and reliable crop yield prediction.

REFERENCES

[1] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritschi, "Soybean yield prediction from UAV using multimodal data fusion and deep learning," Remote Sens. Environ., vol. 237, p. 111599, 2020, doi: 10.1016/j.rse.2019.111599.

[2] A. Avneri et al., "UAS-based imaging for prediction of chickpea crop biophysical parameters and yield," Comput. Electron. Agric., vol. 205, p. 107581, 2023, doi: 10.1016/j.compag.2022.107581.

[3] M. Meroni, F. Waldner, L. Seguini, H. Kerdiles, and F. Rembold, "Yield forecasting with machine learning and small data: What gains for grains?," Agric. For. Meteorol., vol. 308–309, p. 108555, 2021, doi: 10.1016/j.agrformet.2021.108555.

[4] E. Y. Obsie, H. Qu, and F. Drummond, "Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms," Comput. Electron. Agric., vol. 178, p. 105778, 2020, doi: 10.1016/j.compag.2020.105778.

[5] M. A. Lee, A. Monteiro, A. Barclay, J. Marcar, M. Miteva-Neagu, and J. Parker, "A framework for predicting soft-fruit yields and phenology using embedded, networked microsensors, coupled weather models and machine-learning techniques," Comput. Electron. Agric., vol. 168, p. 105103, 2020, doi: 10.1016/j.compag.2019.105103.

[6] V. Sagan et al., "Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning," ISPRS J. Photogramm. Remote Sens., vol. 174, pp. 265–281, 2021, doi: 10.1016/j.isprsjprs.2021.02.008.

[7] P. Abbaszadeh, K. Gavahi, A. Alipour, P. Deb, and H. Moradkhani, "Bayesian Multi-modeling of Deep Neural Nets for Probabilistic Crop Yield Prediction," Agric. For. Meteorol., vol. 314, p. 108773, 2022, doi: 10.1016/j.agrformet.2021.108773.

[8] M. Cheng et al., "Combining multi-indicators with machine-learning algorithms for maize yield early prediction at the county-level in China," Agric. For. Meteorol., vol. 323, p. 109057, 2022, doi: 10.1016/j.agrformet.2022.109057.

[9] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, and S. Shearer, "Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield," Comput. Electron. Agric., vol. 153, pp. 213–225, 2018, doi: 10.1016/j.compag.2018.07.016.

[10] M. Qiao et al., "Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks," Int. J. Appl. Earth Obs. Geoinf., vol. 102, p. 102436, 2021, doi: 10.1016/j.jag.2021.102436.

[11] M. Faheem, M. A. Ngadi, and V. C. Gungor, "Energy efficient multi-objective evolutionary routing scheme for reliable data gathering in Internet of underwater acoustic sensor networks," Ad Hoc Networks, vol. 93, p. 101912, 2019, doi: 10.1016/j.adhoc.2019.101912.

[12] D. Câmara, "Swarm Intelligence (SI)," in Bio-inspired Networking, Elsevier, 2015, pp. 81–102. doi: 10.1016/b978-1-78548-021-8.50004-1.

[13] R. Jaganathan and R. Vadivel, "Intelligent Fish Swarm Inspired Protocol (IFSIP) for Dynamic Ideal Routing in Cognitive Radio Ad-Hoc Networks," Int. J. Comput. Digit. Syst., vol. 10, no. 1, pp. 1063–1074, 2021, doi: 10.12785/ijcds/100196.

[14] J. Ramkumar and R. Vadivel, "Improved Wolf prey inspired protocol for routing in cognitive radio Ad Hoc networks," Int. J. Comput. Networks Appl., vol. 7, no. 5, pp. 126–136, 2020, doi: 10.22247/ijcna/2020/202977.

[15] J. Ramkumar, R. Vadivel, and B. Narasimhan, "Constrained Cuckoo Search Optimization Based Protocol for Routing in Cloud Network," Int. J. Comput. Networks Appl., vol. 8, no. 6, pp. 795–803, 2021, doi: 10.22247/ijcna/2021/210727.

[16] J. Ramkumar and R. Vadivel, "CSIP—cuckoo search inspired protocol for routing in cognitive radio ad hoc networks," in Advances in Intelligent Systems and Computing, 2017, vol. 556, pp. 145–153. doi: 10.1007/978-981-10-3874-7_14.

[17] J. Ramkumar and R. Vadivel, "Whale optimization routing protocol for minimizing energy consumption in cognitive radio wireless sensor network," Int. J. Comput. Networks Appl., vol. 8, no. 4, pp. 455–464, 2021, doi: 10.22247/ijcna/2021/209711.

[18] J. Ramkumar and R. Vadivel, "Multi-Adaptive Routing Protocol for Internet of Things based Ad-hoc Networks," Wirel. Pers. Commun., vol. 120, no. 2, pp. 887–909, Apr. 2021, doi: 10.1007/s11277-021-08495-z.

[19] R. Jaganathan and V. Ramasamy, "Performance modeling of bio-inspired routing protocols in Cognitive Radio Ad Hoc Network to reduce end-to-end delay," Int. J. Intell. Eng. Syst., vol. 12, no. 1, pp. 221–231, 2019, doi: 10.22266/IJIES2019.0228.22.

[20] P. Menakadevi and J. Ramkumar, "Robust Optimization Based Extreme Learning Machine for Sentiment Analysis in Big Data," 2022 Int. Conf. Adv. Comput. Technol. Appl. ICACTA 2022, pp. 1–5, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9753203.

[21] J. Ramkumar, C. Kumuthini, B. Narasimhan, and S. Boopalan, "Energy Consumption Minimization in Cognitive Radio Mobile Ad-Hoc Networks using Enriched Ad-hoc On-demand Distance Vector Protocol," 2022 Int. Conf. Adv. Comput. Technol. Appl. ICACTA 2022, pp. 1–6, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9752899.

[22] R. J, "Meticulous Elephant Herding Optimization based Protocol for Detecting Intrusions in Cognitive Radio Ad Hoc Networks," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 8, pp. 4548–4554, 2020, doi: 10.30534/ijeter/2020/82882020.

[23] J. Ramkumar, "Bee inspired secured protocol for routing in cognitive radio ad hoc networks," Indian J. Sci. Technol., vol. 13, no. 30, pp. 2159–2169, 2020, doi: 10.17485/ijst/v13i30.1152.

[24] J. Ramkumar and R. Vadivel, "Improved frog leap inspired protocol (IFLIP) – for routing in cognitive radio ad hoc networks (CRAHN)," World J. Eng., vol. 15, no. 2, pp. 306–311, 2018, doi: 10.1108/WJE-08-2017-0260.

[25] R. Jaganathan, V. Ramasamy, L. Mani, and N. Balakrishnan, "Diligence Eagle Optimization Protocol for Secure Routing (DEOPSR) in Cloud-Based Wireless Sensor Network," 2022, doi: 10.21203/rs.3.rs-1759040/v1.

[26] K. Gavahi, P. Abbaszadeh, and H. Moradkhani, "DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting," Expert Syst. Appl., vol. 184, p. 115511, 2021, doi: 10.1016/j.eswa.2021.115511.

[27] J. Cao, Z. Zhang, F. Tao, Y. Chen, X. Luo, and J. Xie, "Forecasting global crop yields based on El Nino Southern Oscillation early signals," Agric. Syst., vol. 205, p. 103564, 2023, doi: 10.1016/j.agsy.2022.103564.

[28] S. Fei, Z. Chen, L. Li, Y. Ma, and Y. Xiao, "Bayesian model averaging to improve the yield prediction in wheat breeding trials," Agric. For. Meteorol., vol. 328, p. 109237, 2023, doi: 10.1016/j.agrformet.2022.109237.

[29] A. A. Correndo et al., "Unraveling uncertainty drivers of the maize yield response to nitrogen: A Bayesian and machine learning approach," Agric. For. Meteorol., vol. 311, p. 108668, 2021, doi: 10.1016/j.agrformet.2021.108668.

[30] M. D. Johnson, W. W. Hsieh, A. J. Cannon, A. Davidson, and F. Bédard, "Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods," Agric. For. Meteorol., vol. 218–219, pp. 74–84, 2016, doi: 10.1016/j.agrformet.2015.11.003.

[31] A. Houshmandfar et al., "Machine learning produces higher prediction accuracy than the Jarvis-type model of climatic control on stomatal conductance in a dryland wheat agro-ecosystem," Agric. For. Meteorol., vol. 304–305, p. 108423, 2021, doi: 10.1016/j.agrformet.2021.108423.

[32] B. Zhu, Y. Feng, D. Gong, S. Jiang, L. Zhao, and N. Cui, "Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data," Comput. Electron. Agric., vol. 173, p. 105430, 2020, doi: 10.1016/j.compag.2020.105430.

[33] O. Bazrafshan, M. Ehteram, Z. G. Moshizi, and S. Jamshidi, "Evaluation and uncertainty assessment of wheat yield prediction by multilayer perceptron model with bayesian and copula bayesian approaches," Agric. Water Manag., vol. 273, p. 107881, 2022, doi: 10.1016/j.agwat.2022.107881.

[34] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324/METRICS.

[35] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." Proceedings of the fifth annual workshop on Computational learning theory. 1992.