



*Research article*

## **Analysis and evaluate of agricultural resources using data analytic methods**

**Min Tang\***

School of Marxism, Xi'an Jiaotong University, Xi'an 710049, China

\* **Correspondence:** Email: tangminxjtu@163.com.

**Abstract:** In the agricultural sector, farmers and agribusiness are confronted with a multitude of complex choices every day. These selections are influenced by multiple variables that significantly affect their outcomes. The primary source of revenue for a good deal of individuals is derived from the agricultural sector. The provision of precise and punctual predictions on crop yields has significant importance in facilitating informed investment choices and shaping agricultural policies. One of the challenges encountered is the presence of old or incomplete data about the accessibility of resources. This represents a significant obstacle in accurately ascertaining the present state of affairs. The process of evaluating becomes complex as a result of the diverse range of soil conditions and climatic factors. This research introduces a novel approach called Enhanced Gravitational Search Optimized based Gated Recurrent Unit (EGSO-GRU) for the purpose of calculating crop production. The dataset was first gathered and pre-processed using a normalization method. Enhanced independent component analyses (EICA) have been employed for the purpose of extracting features. To determine the suggest method achievement with regard to accuracy (95.89%), specificity (92.4%), MSE (0.071), RMSE (0.210) and MAE (0.199). The proposed method achieved greater crop prediction accuracy, outperforming the majority of the existing models. The necessity of this progress is vital to the successful operation of crops. The concept signifies a technological advancement aimed at optimizing agricultural resources, hence fostering enhanced productivity and long-term sustainability within the farming industry.

**Keywords:** agriculture research; data analytic method; crop estimation; enhanced gravitational search optimized based gated recurrent unit (EGSO-GRU); enhanced independent component analyses

---

## 1. Introduction

### 1.1. Overview of horticulture

China has become one of the world's leading manufacturers of agricultural products. Gardening is China's most important industry. The expansion of farming is determined by a multitude of economic and environmental factors since gardening represents an atypical industry. Essentially, Andhra Pradesh, with a farming-based economy, provides around 29% of the GDP of the nation than 17% nationwide. Occasional guidance to farmers on improved agricultural techniques or breakthroughs in the factors influencing harvest development might help the state's agricultural business [1]. The global population is deeply worried about the progressively worsening situation of malnutrition. The demand for food to support the world's increasing population has increased over the last several decades despite major improvements in crop production. The shortage of available soil and irrigation resources, the effects of weather evolution and the desire to advance organic farming methods are expected to increase the constraints on current agricultural food production processes [2].

### 1.2. IOT-enable resource management

In farming environments, greater irrigation inspection, chemical control, water supply assessment and fertilizer controls are indicators of optimal resource administration. Past studies have demonstrated that improved use of resources results in significant cost savings. Internet of Things connectivity via Long Range Wide Area Network (LoRaWAN) and other relevant systems may improve resource effectiveness and optimization in agricultural fields and greenhouse cultivation, reduce the cost of production by removing unnecessary human activity and reduce buyer price [3]. The cost reductions were underlined by the decrease in crops caused by climate change and warming temperatures, deforestation and a shortage of agricultural land due to growing cities. According to Environmental Protection, climate change may result in a boost of \$11 billion in chemical expenditure by US farmers. The global expenditures are immeasurable because of increased worries about environmental toxicity, negative effects on human wellness and permanent efficacy [4].

### 1.3. Sustainable food security

Farming materials are critical to global food security and social prosperity. These assets include farmland, water availability, fertile soil, variation in genes in animal and plant life, modern farming methods and skilled labor. Food security, economic stability and environmental preservation depend on the effective management and sustainable use of these resources. As the world's population rises, protecting and maximizing agricultural resources is more important than ever, promoting resilient and fruitful systems that can fulfill the world's growing food demand while preserving its ecosystems [5]. Global warming, extreme weather and natural disasters may all have a severe influence on harvests and animal productivity, leading to food insecurity and economic instability. Overuse of land and water supplies may exacerbate agricultural business issues, resulting in the degradation of soil, loss of biodiversity and shortages of water [6]. Furthermore, heavy dependence on chemical ingredients like pesticides and fertilizer may degrade the planet and pose dangers to consumers and the agricultural workforce.

#### 1.4. Environmental challenges

The provisions of agriculture encompass a diverse variety of organic substances required for cultivating foodstuffs, fiber and fuel. The natural assets that enable food production across the globe vary from fertile soils to abundant water sources for cultivation. The variety of the environment, which includes animals, microorganisms and plants, is critical to sustaining ecological equilibrium and adaptation. Additionally, contemporary agricultural techniques, innovative technology and sustainable management practices remain to successfully use these resources, ensuring food security and financial security for citizens all over the globe. Humans must use these distinctive assets responsibly if they are to secure the long-term viability of farming and feeding the world [7].

#### 1.5. Objective of the study

The recommended method creates a crop yield computation using an EGSO-GRU. Statistics on crop production show the level of agricultural output in a certain region. The initial collection and pre-processing of the dataset uses the normalization procedure.

#### 1.6. Contribution of the paper

- Using the normalization procedure, the dataset was pre-processed.
- Researchers identify the feature used in the Enhanced Independent Component Analysis (EICA).
- The study has created an enhanced Gravitational Search Optimized Gated Recurrent Unit (EGSO-GRU) to forecast crops.

The remainder of this report is divided into the following sections: The literature survey is provided in Section 2. The method is described in Section 3. The experimental findings and discussion are described in Section 4, and Section 5 presents both the conclusion and suggestion for further research.

## 2. Related works

Research in many areas must include a literature survey, commonly called a literature review or systematic review. It entails a thorough review and analysis of the body of research, which is shown in Table 1.

**Table 1.** Literature survey.

References	Objectives	Summary of findings
[8]	The research used a combination of “Pareto analysis, fuzzy theory, Total Interpretative Structural Modeling (TISM)” and MICMAC analysis to categorize, evaluate and examine the hierarchical connections among the Driving factors for agriculture 4.0 in a setting of expanding economic activity.	The study’s outcomes, the three most important factors influencing the adoption of Agriculture 4.0 are “Automation & Robotics to lessen the usage of manual labor,” “Augmented reality (AR) for training new farmers and guiding tool evaluation,” and “Providing adequate financial incentives to impoverished farmers for building modern storage facilities.”

*Continued on next page*

References	Objectives	Summary of findings
[9]	The paper proposed supporting data analysis activities in agricultural environments, a multi-domain data fusion technique. They also include all the data sources gathered and used to create.	The combined information sources were assessed using a multi-label learning approach's early exploratory analysis. Finally, a case study in agricultural crop production describes the data fusion technique.
[10]	The research examines the mediating role of data analytics capabilities using the dynamic capability theory and looks at how the use of big data analysis impacts the effectiveness of findings.	The empirical findings demonstrated a favorable correlation between the use of big data analytics and the quality of decisions made, with data analytics skills acting as a mediating factor in this connection.
[11]	The paper suggested that important areas connected to agricultural drones include machine learning, deep learning, the Internet of Things, agricultural precision, mapping and in-depth perception are all examples of artificial intelligence.	The literature's co-citation analysis reveals six main research clusters. One of the initial initiatives is to gather drone research in agriculture and lay out possible future investigations.
[12]	The study of vegetable sources has encouraged the screening of raw materials for new antioxidants by the desire to replace synthetic food antioxidants with natural ones. Their extraction from cheap or leftover agricultural industry supplies is given special consideration.	The trivalent iron reduction method, the radical ABTS assay and the DPPH test were employed to assess the antioxidant activity of the lignin under study. Researchers found that the antioxidants in lignin may reduce oxidants and balance out free radicals.
[13]	The paper examined the viability of using recycled water for irrigation purposes in California utilizing a combined geographic "Analytical Hierarchy Process (AHP) and Multi-Criteria Decision Analysis (MCDA)."	The outcomes also showed how closely connected the agricultural regions are to the geographical distribution of places appropriate for reusing recovered water.
[14]	The paper used the indicators water resources use (WRU), efficiency (WRE) and productivity (WRP) for measuring overall water exploitation, effective utilization rate and production capacity, respectively.	The outcomes also showed how closely connected the agricultural regions are to the geographical distribution of places appropriate for reusing recovered water.
[15]	The paper examined mining methods and techniques and how they apply to digital agriculture from a big data perspective.	Then, it offered a broad discussion on the application of big data analytics to the field of digital agriculture after examining each category of the categorization through various current works and demonstrating their applied approaches.
[16]	The study employed the water quality of surface waters utilized for agricultural irrigation, "two water quality indices (AHP-IWQI and DEA-IWQI) based on Data Envelopment Analysis and Analytic Hierarchy Process were developed."	The DEA-IWQI has a higher accuracy ratio than the AHP-IWQI, according to findings of comparisons between predictions made using AHP- and DEA-based indices.

*Continued on next page*

References	Objectives	Summary of findings
[17]	Big data analytics has the potential to improve the agriculture sector significantly. Agriculturalists rely primarily on traditional procedures to choose the best crop to plant and how to grow it effectively and profitably.	Integrating big data with the industry might drastically alter how agriculture is conducted. Using big data analytics properly, it is possible to comprehend the crop production method.
[18]	The paper suggested a framework for assessing the effectiveness of the assessment system that combines the “TOPSIS (technology for order performance by similarity to ideal solution)” and entropy methods.	Assessing and analyzing agricultural extension services may produce an efficient analytical tool for sustainable agriculture.
[19]	The study examined and derived the elements influencing the growth of agricultural goods as its research object.	Using a fuzzy comprehensive assessment technique and the analytical hierarchy process, analysis and induction of elements impacting the growth of agricultural products, collection of pertinent data and comparison analysis to validate the evaluation.
[20]	This study proposed a method for evaluating the area of land used for peri-urban agriculture, the “Analytical Hierarchical Process (AHP) and geographic information system (GIS).”	The suitability map was produced by linearly integrating five geographical layers with corresponding weights. Further scales included highly appropriate land for peri-urban agriculture in three categories: Medium, low and unsuitable.
[21]	In-depth information about the management of different farm-related activities using a variety of technologies, including “IoT, UAVs, IoT, Big Data Analytics, Deep Learning Techniques and Machine Learning” techniques, is provided by this study.	These subjects include a general overview of relevant technology, their applications and recent case studies and research projects show that to employ these technologies in Agriculture 4.0.
[22]	They offered an organized look at several big GIS analytics (BGA) in agricultural applications. The extent of GIS and analytics applications in agriculture is the two main areas into which the chosen papers are divided.	The study’s recommended conclusion is a BGA framework for the agricultural supply chain.
[23]	Several IoT advantages and difficulties have been mentioned in this research. They outline the IoT ecosystem and how DA and IoT work together to enable smart agriculture.	The Internet of Things (IoT) and data analytics (DA) are used to improve operational effectiveness and productivity in the agriculture industry.
[24]	This article examined many barriers to developing the agricultural supply chain and different industry 4.0 solutions to overcome problems.	This study demonstrated that incorporating technical developments into agriculture may improve supply chain management. Additionally, it shows how the massive amounts of data produced may address these escalating problems in the agricultural supply chain.

*Continued on next page*

References	Objectives	Summary of findings
[25]	Multi-criteria decision-making (MCDM) procedures are effective in various complicated decision-making situations in the real world, including agricultural difficulties.	The models employed, the data sources used and the precision attained over the previous few years using various performance metrics.
[26]	A surrogate-based optimization methodology was presented for massive amounts and high-24 precision landscape management optimization. To develop a 26 surrogate of the DayCent biogeochemical simulation model, an artificial neural network was utilized.	The outcomes demonstrate that more than 99% of the variances 30 in DayCent's simulated outputs were caught by the surrogate. In comparison to the "business-as-usual" scenario, farm-level optimization raised farm income by 83%
[27]	The study suggested a transfer learning-based Deep Convolutional Neural Network (DCNN) method for the precise identification and categorization of rice leaf disease. One of the updated suggested approach's transfer learning techniques was based on VGG19.	The overall average precision of the non-normalized augmented sample is 96.08%. The associated F1-score, recall, accuracy and specificity were, in that order, 0.9621, 0.9627, 0.9931 and 0.9626.
[28]	The study focuses on enhancing overall efficacy and efficiency by using cutting-edge technology and large-scale data analysis. It was discovered that a number of technologies, including cloud computing, image processing, Convolutional neural networks (CNN), wireless sensor networks, Internet of Things (IoT) and remote sensing, could be applied.	These tools could help business owners save money and time as they offer a dependable service for their clients. In general, these contemporary technologies have the potential to support in the long run and guarantee the industry's sustainable expansion.
[29]	The study produced an enormous amount of food to sustain the world's expanding population. It has greatly improved human health and welfare. However, it also places significant strain on the surroundings and natural resources.	To enhance soil microbial diversity, better soil nutrients, decreased evolution in the related Phytophthora infestans pathogen and higher crop resilience and pathogen reductions. These complementary advantages, when combined, offer a strong chance to realize sustainable agriculture.
[30]	The study utilized the soil, carbon, nitrogen and aquatic footprints to calculate the greenhouse gas (GHG) releases and resource requirements of agriculture. The amount that efficient growth could allow humanity to satisfy its need for food products was currently determined.	To satisfy potential food demand, production efficiency gains simply could not be enough if the environmental cost of food production failed to eventually rise.

### 2.1. Problem statement

The preceding statement emphasizes the importance of cropping yield estimation in agricultural planning and management, emphasizing the critical role of modern technologies and data-driven methods in forecasting possible crop yields. Despite the growing significance of precise projections

for optimal use of resources and ethical procedures, there remains a significant research gap that requires further analysis and investigation. One critical area that requires focus is the integration and enhancement of various sources of data. Although the assertion recognizes the significance of factors such as climate patterns, quality of soil, crop condition and historical data, more research is needed to determine how these elements may be efficiently linked and studied in real-time. Due to the dynamic and interlinked aspect of these influences, research may concentrate on creating more complex methods or ML methods that can exploit all of them.

Additionally, the time-related element of yield estimation necessitates further investigation. Present methods may be insufficient to address the issues given by rapid modifications to natural circumstances of climatic changes or unexpected catastrophes. Developing adaptable models that can react rapidly to changing circumstances and unanticipated interruptions may considerably improve yield expectation quality and dependability. There is also a study deficit in comprehending the socioeconomic ramifications of precise crop output estimates. While the assertion mentions the advantages of allocation of resources optimization and legislators, further research is needed to determine how exact yield projections might assist farmers' earnings, local economies and general food security. In conclusion, the research gap is defined by the need for additional methods to integrate multiple data sources, construct adaptive models and investigate the broader socioeconomic consequences of accurate crop output predictions.

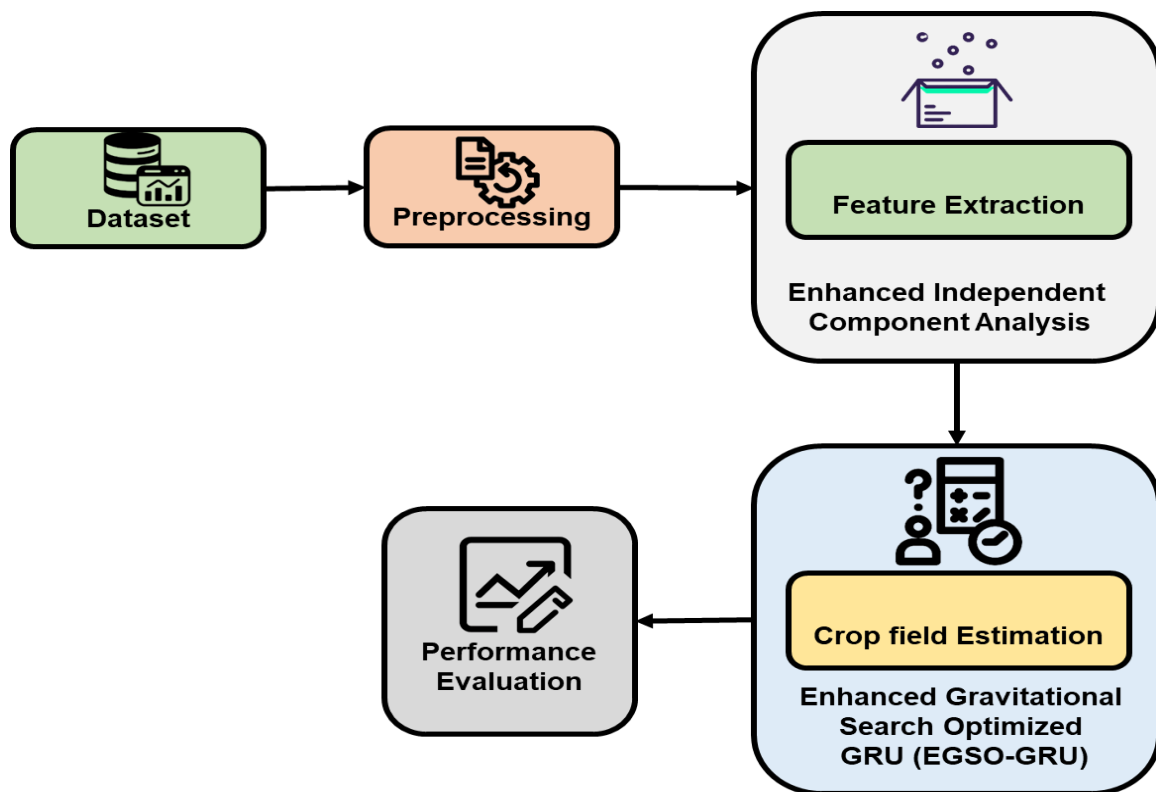
### 3. Proposed method

For farmers and policymakers to make informed decisions, crop yield prediction is a critical component of contemporary agriculture. Crop yields may be accurately predicted using cutting-edge technology like remote sensing, satellite photography and data analytics. These techniques analyze a variety of variables, including weather patterns, soil health, historical data and crop growth phases, to offer insightful information. In addition to increasing the effectiveness of agricultural practices, the data-driven approach also helps in resource management, providing food security and sustainability for a growing global population. Crop yield estimation is a critical issue in agriculture since it aids farmers, policymakers and researchers in decision-making about crop management, food security and resource allocation. Various techniques exist for estimating agricultural output, and technological improvements have created more precise and effective systems. We cover current cutting-edge study on implementing artificial intelligence (AI) through adaptive greenhouses with a goal to increase harvest rates, water and nitrogen intake effectiveness, illness and pest control and sustainable agriculture. Automation devices for using chemicals, irrigation produce and harvesting them, along with bio-inspired techniques for streamlining greenhouse procedures, handling energy, equipment route organizing and managing unmanned aerial vehicles (UAVs), resolving scheduling problems and evaluating images for indications of illness and pests, were between the most studied issues [31]. since neural networks approximated brain function, bio-inspired ANN algorithms fared better. Whereas, GA algorithms were preferred for agricultural organizing, applying pesticides, equipment route optimization and other simulations. AFSOA, firefly, ACOA, ABCOA and Cuckoo methods have enhanced land utilization organizing, yield optimization, agriculture resource utilization (waters, ground, crop chemicals and nutrients), along with supply chain processes (afterwards managing of fruits and veggies) [32].

When analyzing the agricultural supply chain, it is crucial to assess the effects of service aspects

and the adoption of buy-online-pick-up-in-store (BOPIS) methods within an omnichannel framework. Service variables involve several concerns, such as the velocity, dependability and general superiority of services offered throughout the supply chain. This assessment encompasses the effectiveness of a delivery process, accuracy in fulfilling the needs of crops and the quality of customer service. Additionally, BOPIS initiatives allow farmers and organizations to buy crops via the Internet and pick goods up at approved sites. These techniques depend on an omnichannel platform that provides a consistent experience across internet portals, physical shops and mobile apps. Our objective is to comprehend how these components affect the agricultural supply chain to optimize operations, improve customer happiness and streamline agricultural product distribution.

Agriculture is vulnerable to weather, pests and market variations. The inventory model may modify and optimize to these uncertainties because to the ANN's pattern recognition. It can dynamically adjust crops and harvesting schedules depending on real-time weather and market needs, improving agricultural resilience. The influence of inflation on agriculture's input costs and price dynamics is also important. The integrated ANN, trained to account for inflation, helps the inventory model allocate resources and price and choose crops. This adaptability helps farmers and agricultural enterprises manage inflation and make smart decisions to stay profitable. Figure 1 shows a recommended approach for estimating agricultural production that blends traditional and modern techniques:



**Figure 1.** Architecture of the proposed method.

#### a) Dataset

A dataset is the starting point for understanding and improving agricultural productivity in crop yield analysis. This extensive data set includes various factors, such as seasonal climate, soil properties, crop varieties and associated yields. A sizable generic crop dataset with agricultural parameters feeds



the model. As the feature dataset, another dataset was used. A website called kaggle.com was where the datasets were gathered. The crop dataset takes up 7841 kb of space. This dataset's prediction factors include temperature, rainfall, pH level, relative humidity and geographic location. These crops include wheat, rice, maize, millet, peas, pigeon peas, sugarcane and green gram. Every forecast parameter for a specific yield has a range of available values. For instance, if the crop was wheat, any value within the scope of wheat values available in the dataset can be used for the prediction parameters. It holds for all of the crops included in the dataset.

#### b) Pre-processing

In crop yield analysis, the pre-processing stage is essential for drawing important conclusions from unprocessed agricultural data. Data collection and cleaning are included in this first stage, which entails gathering and standardizing data on weather patterns, soil properties and historical yield records. First, it must remove all of the missing values from the input dataset since it has some missing values. All the missing values in the input dataset are removed using a pre-processing procedure. The proposed approach employs normalization and variable mean techniques.

There are several approaches to normalizing a dataset. The most effective normalization in this case is Minmax. The normalizing technique converts  $\sigma_{max}$  to  $\sigma_{min}$ , which lies between [A, B]. Equation (1) provides the mathematical formula for it;

$$\sigma^* = \frac{\sigma - \sigma_{min}}{\sigma_{max} - \sigma_{min}} \quad (1)$$

In Eq (1),  $\sigma_{max}$  represent the Maximum value and  $\sigma_{min}$  represent the minimum value and  $\sigma^*$  represent the normalized value. All attributes in the data are given the same weight after normalization. Q stands for the lowest value and B for the highest value.

#### c) Feature extraction using Enhanced independent component analyses (EICA)

A multivariate dataset can be divided into independent non-Gaussian components using the statistical and computational method of enhanced independent component analysis (EICA). Crop yield forecast considers several variables, including weather, soil composition, farming techniques and the use of fertilizers and pesticides. Regression models, machine learning algorithms and crop growth models are examples of conventional methods for predicting crop yields, including pertinent variables to calculate a crop's prospective yield. The removal of tiny information and the reduction of spectral differences between pixels in the same class were both accomplished through image denoising. This smoothing process also made distinguishing between pixels belonging to various courses harder, impacting how well the categorization works. ICA is a great solution to this problem since it can emphasize spectral differences between pixels of different classes while extracting the most important characteristics.

The data matrix S, which has N observations and M variables, encodes the basic equation driving ICA-based sparse representation. In the setting of image processing the N represents the total number of pixels of an image, and M is the value of every pixel's characteristic.

$$B = U'G \quad (2)$$

P values represent the primary components created by averaging the initial sample vectors. P comprises parts intended to preserve more variation than in the first parts. The covariance matrix C defines W.

$$U = AE^{-\frac{1}{2}} \quad (3)$$

In this scenario, A is a diagonal matrix with C's decreasing eigenvalues, while E contains C's rising eigenvectors. It is reasonable to suppose that since  $G_M - \mu$  is the  $n^{th}$  column in the M by N matrix B.

$$P = [G_1 - \mu, \dots, G_M - \mu] \quad (4)$$

Inputting the data into the  $\mu = (1/M)(G_1 + \dots + G_M)$  formula yields the average vector. One may find the covariance matrix C of size  $M \times M$  using the following formula.

$$V = \frac{1}{M-1} PP^D \quad (5)$$

The scaling factor W for the E vector matrix is inferred from Eq (3) such that each independent variable  $P_m$  variance is 1. The first L significant components are kept, which reduces the dimensions of the input data S. The proposed technique, the trained framework, is represented by the formula ICA S; L, where L is the number of components to hold from the data matrix S. Basic theory and ICA equations are explained in this section.

d) *Crop yield estimation of Enhanced Gravitational Search Optimized based Gated Recurrent Unit (EGSO-GRU)*

### 3.1. Enhanced Gravitational Search Optimization (EGSO)

In contrast to other recognized population-based optimization methods, GSA's search agents are a group of masses that interact according to Newton's principles of gravity and motion. The agents' performance is seen in GSA as an object whose mass determines its performance. All items are attracted to one another by gravity, which generates an overall motion of all objects in the direction of the objects with heavier masses.

The location of the  $i$ th agent, assuming  $N$  agents (masses), A for B. Newton's theory of gravity defines the force coming from the  $j$ th mass toward the  $i$ th mass as follows:

$$E_{ji}^c(s) = H_{(s)} \frac{N_j(s) \times N_i(s)}{\|W_j(s), W_i(s)\|_2} (w_j^c(s) - w_i^c(s)) \quad (6)$$

The gravitational constant at time  $t$  is  $G(t)$ , and  $N_j$  and  $N_i$  are the masses of the agents.  $N_j$  is computed using a fitness  $N_j(s) = \frac{fit_j(s) - worst(s)}{best(s) - worst(s)}$  comparison;  $best(t)$  represents the greatest capability of all agents, and  $worst(t)$  represents the poorest suitability of every agent. The total of the randomly weighted pressures applied by the other agents, for the  $i$ th agent,

$$E_j^c(s) = \sum_{i \neq j} rand_i E_{ji}^c(s) \quad (7)$$

The  $i^{th}$  agent's acceleration is determined using the law of motion as follows:

$$b_j^c(s) = \frac{E_j^c(s)}{N_i(s)} \quad (8)$$

The following equations may then be used to explain the search method for this notion.

$$u_j^c(s+1) = rand_i \times u_j^c(s) + b_j^c(s) \quad (9)$$

$$w_j^c(s+1) = w_j^c(s) + u_j^c(s+1) \quad (10)$$

A represents the location of the  $i$ th agent in the  $d$ th dimension in Eqs (6) and (7), whereas B denotes velocity and C denotes acceleration.

$$H(s) = H_0 \cdot \exp\left(-\beta \cdot \frac{s}{max-s}\right) \quad (11)$$

It must be noted that the performance of GSA is influenced by the gravitational constant( $s$ ), which is specified as a function of time  $t$ .  $H_0$  is the starting value,  $s$  is a constant,  $t$  is the current iterations and  $max\_t$  represents the maximum iterations.

The parameters that determine GSA's performance are N-fold population growth, starting gravitational constant  $H_0$ ,  $max\_t$  and constant  $b$ . It has been demonstrated that GSA is highly effective in solving a variety of nonlinear functions. GSA, an effective optimization method, can resolve various optimization issues. We use GSA to determine the parameters of a difficult nonlinear system.

### 3.2. Improvements in GSA

That draws inspiration from the law of global gravitation instead of algorithms replicating animal social behavior. While GSA is memory-less and considers the agents' present positions when updating, the searching operation that other agents have accumulated determines the path of an agent. Where  $p$  best is the  $i$ th particle's best prior position, and  $g$ best is the sum of all particle's best initial positions,  $q_{j1}$  and  $q_{j2}$  are two random variables with values in  $[0, 1]$ ,  $d_1$  and  $d_2$  are positive constants, and  $w$  is the inertia weight.

$$u_j^c(s+1) = x(s)w_j^c(s) + d_1q_{j1}(obest_j^c - w_j^c(s)) + d_1q_{j1}(hbest_j^c - w_j^c(s)) \quad (12)$$

$$w_j^c(s+1) = w_j^c(s) + u_j^c(s+1)w_j^c(s) + d_1q_{j1}(obest_j^c - w_{j++++}^c(s)) \quad (13)$$

The "Improved Gravitational Search Algorithm (IGSA)" uses a cutting-edge movement technique when searching according to gravity's laws and taking cues from social and memory cues. Following is a definition of the IGSA velocity updating equation:

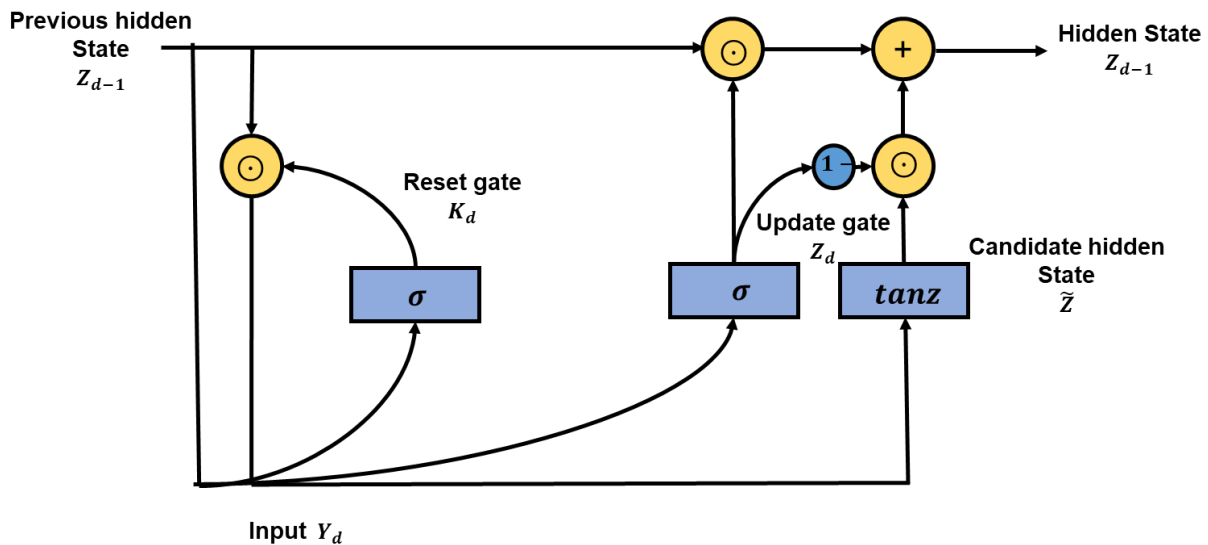
$$u_j^c(s+1) = q_{j1}u_j^c(s) + b_j^c(s) + d_1q_{j2}(obest_j^c - w_j^c(s)) + d_1q_{j3}(hbest^c - w_j^c(s)) \quad (14)$$

where  $d_1$  and  $d$  are variables in the  $[0, 1]$  range and  $q$ ,  $q_{j2}$  and  $q_{j3}$  are random variables. They may balance efficacy about "law of gravity" and "memory and social information" by varying the values of  $d$  and  $d_2$ . It is evident from Eq (14) that a hybrid and widely-used version of GSA is IGSA. When  $d_1$  and  $d_2$  are equivalent, IGSA and GSA equal zero.

### 3.3. Gated Recurrent Unit (GRU)

Estimating crop yields is a crucial undertaking in contemporary agriculture since it enables farmers to allocate resources and manage crops in an educated manner. By employing Enhanced Gravitational Search Optimized based on Gated Recurrent Unit (GRU) models, a unique method has been devised to increase the accuracy of this estimation process. The GRU is a recurrent neural network well-suited for sequential data analysis, and it is used in this cutting-edge method to combine the strength of Enhanced Gravitational Search Optimization, a metaheuristic algorithm inspired by nature. This novel strategy promises to give more accurate and trustworthy crop output estimates by combining the advantages of both approaches, opening the door for increased agricultural productivity and environmentally friendly farming methods.

An alternative to the LSTM neural network is the gated recurrent unit (GRU). To increase the impact of early key data on subsequent training, it can learn to update or discard the hidden state when the sequence is too lengthy. To replace the input and forgetting gates in the LSTM, the GRU utilizes update gates and reset gates to replace output gates. The current time step input and the preceding time step's hidden state provide the information for resetting and updating gates, respectively. Figure 2 displays its flow chart:



**Figure 2.** Modelling of GRU's structure.

The input is  $Y_d$  for a specific time step  $d$ . The preceding time step's concealed state is  $Z_{d-1}$ ,  $\sigma$ . This shows that the following definitions apply to the activation function sigmoid, gate  $Z_d$  and update gate  $K_d$ :

$$K_d = \sigma(Y_d U_{yk} + Z_{d-1} U_{zk} + p_k) \quad (15)$$

$$Z_d = \sigma(Y_d U_{yh} + Z_{d-1} U_{zh} + p_h) \quad (16)$$

$P$  is the paranoid parameter, and  $U$  is the weight matrix that needs to be learned. The candidate

hidden state is then obtained by integrating the reset gate  $K$  and the traditional hidden state updating process during the time step:

$$\tilde{Z} = \text{tanz}(Y_d U_{yh} + (K_d \odot Z_{d-1}) U_{hh} + p_z) \quad (17)$$

$\tilde{Z}$  denotes the potential concealed state, while  $Z$  denotes the hidden state. Only the input and the preceding hidden state  $Z_{d-1}$  are relevant to the candidate's secret state. The sign represents Hadamard Product  $\odot$ . To ensure that the candidate concealed state's value is concerning  $(-1,1)$ , employ the tanh nonlinear activation function in this instance.

The new hidden state  $Z_d$  must then be determined by combining the changes to gate  $H_d$ , and old state  $H_d$  and invisible state  $H_{d-1}$  may generally be equated to  $H_d$ . Combining components among  $H_{d-1}$  and  $H_d$  is an update gate's purpose. Hence, the following definitions apply to a concealed status update:

$$H_d = H_d \odot Z_{d-1} + (1 - H_{d-1}) \odot \tilde{Z} \quad (18)$$

The update gate in GRU aids in capturing the sequence's long-term dependencies, while the reset gate aids in capturing its short-term requirements, as shown in Algorithm 1.

---

**Algorithm 1: (EGSO-GRU)**

---

**Step 1:** Initialize the population randomly

*population = create\_random\_population (population\_size, dimension)*

**Step 2:** Evaluate the fitness of each individual

*fitness = evaluate\_fitness (population)*

**Step 3:** Find the best individual (Global best)

*global\_best = find\_best\_individual (population, fitness)*

**Step 4:** Main optimization loop

*for generation in range (max\_generations):*

**Step 4.1:** Calculate the gravitational force for each individual

*gravitational\_force = calculate\_gravitational\_force (population, fitness, global\_best)*

**Step 4.2:** Update velocities and positions

*population = update\_velocities\_and\_positions (population, gravitational\_force)*

**Step 4.3:** Apply crossover and mutation

*population = apply\_crossover\_and\_mutation  
(population, crossover\_rate, mutation\_rate)*

**Step 4.4:** Evaluate the fitness of each individual

*fitness = evaluate\_fitness (population)*

**Step 4.5:** Update global best if needed

*global\_best\_candidate = find\_best\_individual (population, fitness)  
if global\_best\_candidate is better than global\_best:  
global\_best = global\_best\_candidate*

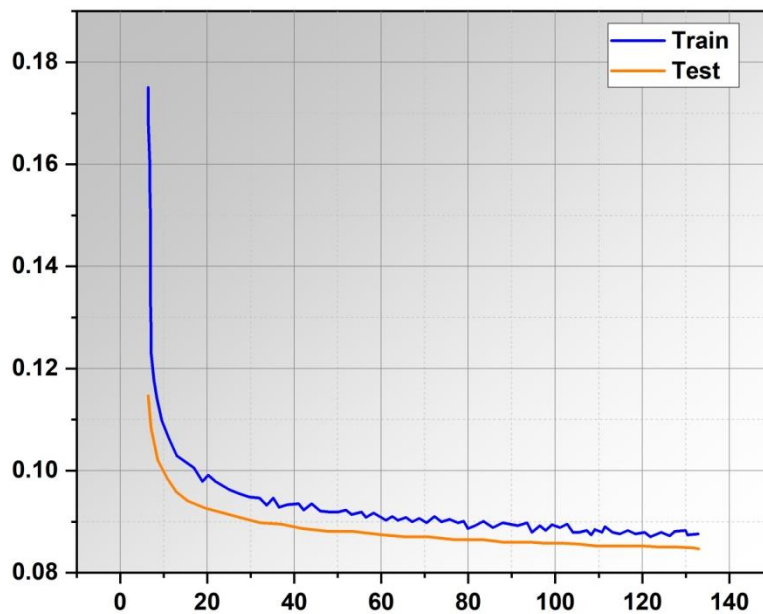
**Step 5:** Return the best individual found

*return global\_best*

---

#### 4. Result and discussion

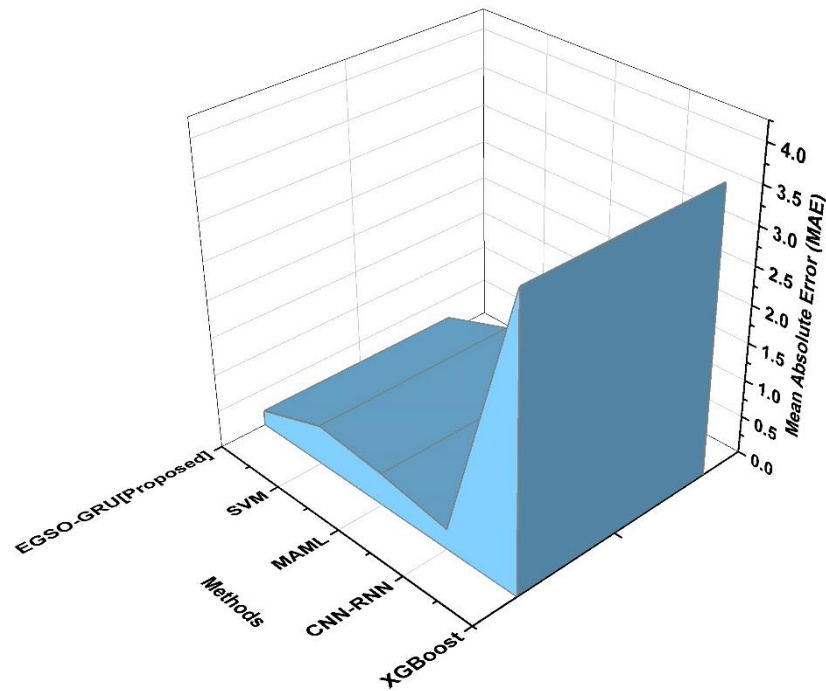
In this paper, Python 3.11 has been used for implementing the enhanced Gravitational Search Optimized Gated Recurrent Unit (EGSO-GRU). A laptop with 32 GB of RAM, an Intel (R) processor and Windows 10 installed was used. The suggested EGS-GRU method was thoroughly evaluated via comparison and assessment of the results. When compared to alternative approaches, the model's contribution to calculating crop yields is an indication that the suggested methodology was successful and effective in Support Vector Machine (SVM), Model Agnostic Meta-Learning (MAML), Convolution Neural Network-Recurrent Neural Network (CNN-RNN) and XGBoost. The graphic shows the accuracy (%) rate applied and the recommended approach for a mean square error (MSE), mean absolute error (MAE), specificity and Root mean square error (RMSE). Feature engineering is a method of selecting the most advantageous features based on predictions and model outcomes. Figure 3 illustrates the model performance that was obtained after 150 epochs of model execution.



**Figure 3.** The efficiency of EGSO-GRU measure.

Mean Absolute Error (MAE): The absolute difference between the anticipated and actual values is used to measure the differences between paired observations representing the same phenomenon.

$$MAE = \sqrt{\frac{1}{n} \sum_{j=1}^n |\hat{z}_j - z_j|} \quad (19)$$



**Figure 4.** Comparison of Mean Absolute Error.

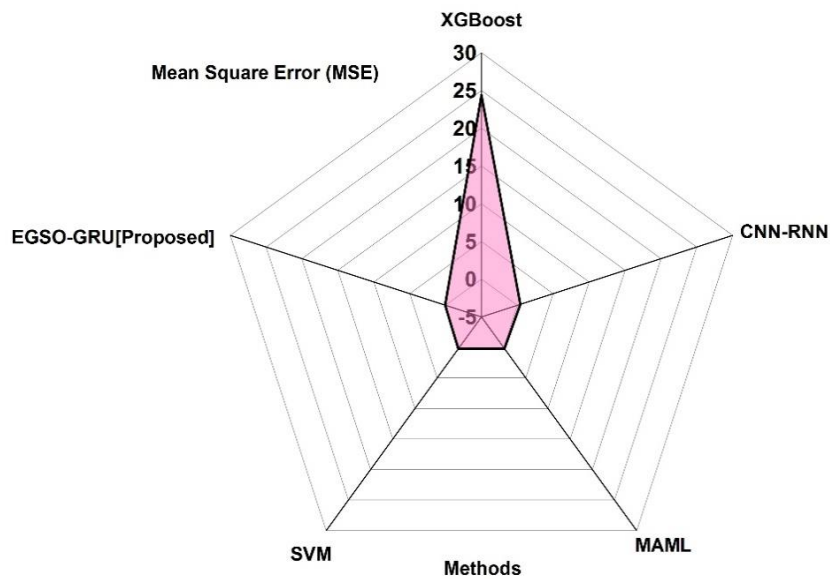
Figure 4 and Table 2 show the Mean absolute error of the suggested approach, and the consumption prediction of the Mean fundamental error in the suggested system and the current systems are indicated. CNN-RNN has acquired 0.269, XGBoost has attained 3.764, MAML has reached 0.426 and SVM 0.314, the proposed method achieves 0.199. It shows that the proposed approach has less value than the existing one.

**Table 2.** Numerical outcomes of Mean Absolute Error.

Method	Mean Absolute Error (MAE)
XG Boost	3.764
CNN-RNN	0.269
MAML	0.426
SVM	0.314
EGSO-GRU [Proposed]	0.199

Mean Square Error (MSE) is calculated as every squared difference between a variable's observed and predicted values divided by the variable's total number of values.

$$MSE = \frac{1}{m} \sum_{j=1}^m (z_j - \hat{z}_j)^2 \quad (20)$$



**Figure 5.** Comparison of Mean Square Error.

Figure 5 and Table 3 show the Mean sequence error of the suggested method and the consumption predicting of Mean sequence error in existing methods, and the suggested method is indicated. XGBoost has attained 24.360, CNN-RNN has acquired 0.415, CNN-LSTM has reached 0.215 and SVM has acquired 0.228. The proposed scheme achieves 0.071. It demonstrates that the suggested strategy is inferior to the current one, so the proposed one is high.

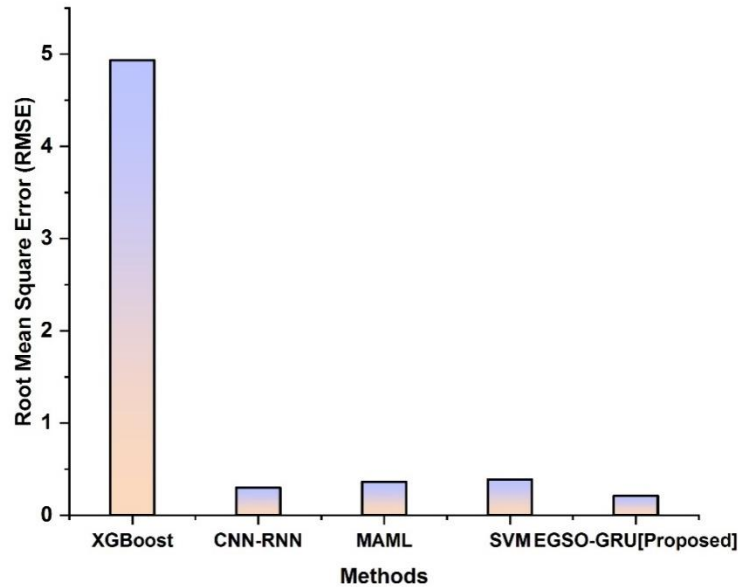
**Table 3.** Numerical outcomes of Mean Square Error.

Methods	Mean Square Error (MSE)
XG Boost	24.360
CNN-RNN	0.415
MAML	0.215
SVM	0.228
EGSO-GRU [Proposed]	0.071

Root Mean Square Error (RMSE) is a rating of the predictor's quality. The RMSE, which represents the standard deviation of the residuals, is the square root of the MSE.

$$MSE = \frac{1}{m} \sum_{j=1}^m (z_j - \hat{z}_j)^2 \quad (21)$$





**Figure 6.** Comparison of Root Mean Square Error.

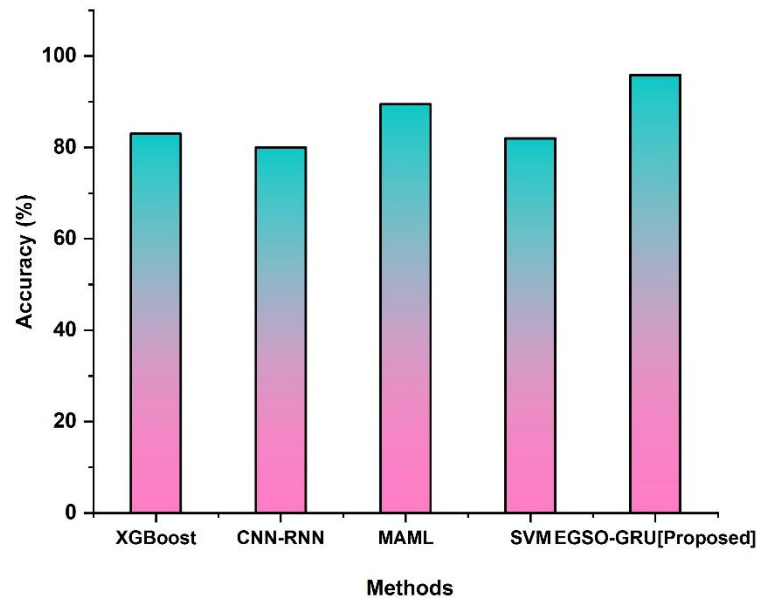
Figure 6 and Table 4 show the root mean square error of the proposed system, and the consumption prediction of source mean square error in the suggested system and the current systems are indicated as 0.210. CNN-RNN has acquired 0.299, XGBoost has attained 4.936%, CNN-LSTM has reached 0.299 and SVM has denoted 0.389 in the proposed method. It shows that the proposed approach has a lower value than the existing one.

**Table 4.** Numerical outcomes of Root Mean Square Error.

Methods	Root Mean Square Error (RMSE)
XG Boost	4.936
CNN-RNN	0.299
MAML	0.364
SVM	0.389
EGSO-GRU [Proposed]	0.210

Accuracy (%) The following formula determines accuracy: the corrected prediction to the total number of predictions.

$$Accuracy = \frac{\text{Correct prediction}}{\text{Total number of predictions}} \quad (22)$$



**Figure 7.** Comparison of accuracy.

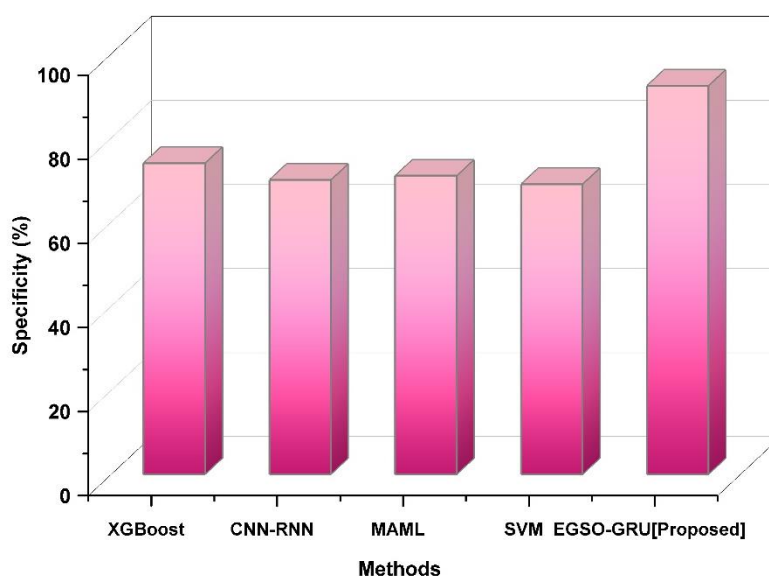
The accuracy metrics for suggested and current strategies are shown in Figure 7 and Table 5. That accurate something reflects how near it is to the real value. The accuracy rate for the suggested methodology employing data analysis methods is 95.89 %, whereas, for the previous process, the accuracy rates for XGBoost, CNN-RNN, SVM and MAML classifiers were 83, 80, 89.5 and 82%, respectively. This proves unequivocally that the recommended approach outperforms the other options in terms of accuracy.

**Table 5.** Numerical outcomes of accuracy.

Methods	Accuracy (%)
XG Boost	83
CNN-RNN	80
MAML	89.5
SVM	82
EGSO-GRU [Proposed]	95.89

Specificity is a metric used in statistics and data analysis to assess how well a binary classification test performed. Out of all real negatives, it determines the percentage of true negatives.

$$Specificity = \frac{TN}{TN+FP} \quad (23)$$



**Figure 8.** Comparison of specificity.

The specificity metrics for current and suggested techniques are shown in Figure 8 and Table 6. 92.4% of the suggested method's specificity is calculated utilizing data analytics. The accuracy values produced by the XGBoost, CNN-RNN, SVM and MAML classifiers in the current technique are 74, 70, 71 and 69%, respectively. This demonstrates unequivocally that the recommended approach's specificity value is higher than the existing methods.

**Table 6.** Numerical outcomes of specificity.

Methods	Specificity (%)
XG Boost	74
CNN-RNN	70
MAML	71
SVM	69
EGS-OGRU [Proposed]	92.4

#### 4.1. Discussion

The existing methods are XGBoost, CNN-RNN, MAML and SVM. A significant amount of data is needed for XGBoost's training. Large-scale, high-quality dataset acquisition in agriculture can be difficult because of things like crop-specific variables, soil variation and weather unpredictability. CNN-RNN models could be difficult to interpret due to their complexity. It can be challenging to understand the model's internal workings and extract useful insights for agricultural decision-making. The tasks that are faced during meta-training and meta-testing are assumed to be similar by MAML. Depending on variables, including varieties of crops, weather, condition of the soil and farming

methods, work in agriculture can differ significantly. It can be difficult to modify MAML to accommodate such task variations. The function of the kernel and its parameters are important factors that affect SVM performance. It can be difficult to choose the right kernel and tuning parameters, and the best options could differ depending on the type of agricultural application. Self-organizing systems can adjust to changing circumstances. This flexibility could be useful in agriculture to deal with changing conditions, including pest outbreaks, soil variances and climatic changes.

#### *4.2. Issue and challenges*

Various issues arise in the investigation and assessment of agricultural resources using data analytics. As agricultural systems produce complex and diverse data, data quality and availability present considerable challenges. Furthermore, the complexity of biological systems, changes in the weather and market dynamics all introduce uncertainty into predictive modeling. It may be difficult to strike a balance between the approaches' scalability and the necessity for detail in analysis. Furthermore, it may be difficult to gather and share data in rural regions due to the digital divide. For the management of agricultural resources in a sustainable manner, these problems must be resolved.

### **5. Conclusions**

We conclude by presenting a novel approach, namely the EGSO-GRU technique, for estimating agricultural yield through the utilization of improved gravitational search optimization. In contrast to the currently employed techniques, the method put forward exhibits the following performance metrics: MAE of (0.199), MSE of (0.071), RMSE of (0.210), Accuracy of (95.89%) and specificity of (92.4%). The findings obtained from the approach suggested have shown a significant level of performance compared to the existing methodologies. One of the challenges encountered pertains to the presence of old or incomplete data regarding the accessibility of resources, hence posing difficulties in accurately assessing the current state of affairs. The process of evaluating becomes increasingly complex as a result of the diverse range of soil conditions and climatic factors.

#### *5.1. Future scope*

The integration of advanced technology like artificial intelligence, precise farming and remote sensing holds the potential to significantly improve agricultural supplies in the future. The optimization of resource distribution, crop management and process of decision-making will be greatly aided by advanced machine learning algorithms, such as deep learning and ensemble approaches.

#### **Use of AI tools declaration**

This research was not used Artificial Intelligence (AI) tools in the creation of this article.

#### **Acknowledgments**

Shaanxi Social Science Foundation Project: Research on the theoretical logic and practical path

of improving Shaanxi agricultural support and protection system in the context of rural revitalization. (NO: 2022XC02).

### Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### References

1. J. E. Koltjes, J. B. Cole, R. Clemmens, R. N. Dilger, L. M. Kramer, J. K. Lunney, A vision for the development and utilization of high-throughput phenotyping and big data analytics in livestock, *Front. Genet.*, **10** (2019), 1197–1202. <https://doi.org/10.3389/fgene.2019.01197>
2. J. Seyedmohammadi, F. Sarmadian, A. A. Jafarzadeh, R. W. McDowell, Development of a model using matter element, AHP, and GIS techniques to assess land suitability for agriculture, *Geoderma*, **352** (2019), 80–95. <https://doi.org/10.1016/j.geoderma.2019.05.046>
3. S. Gokool, M. Mahomed, R. Kunz, A. Clulow, M. Sibanda, V. Naiken, et al., Crop monitoring in smallholder farms using uncrewed aerial vehicles to facilitate precision agriculture practices: a scoping review and bibliometric analysis, *Sustainability*, **15** (2023), 1–18. <https://doi.org/10.3390/su15043557>
4. M. M. Ali, N. Hashim, S. A. Aziz, O. Lasekan, Principles and recent advances in the electronic nose for quality inspection of agricultural and food products, *Trends Food Sci. Technol.*, **99** (2020), 1–10. <https://doi.org/10.1016/j.tifs.2020.02.028>
5. C. Maraveas, D. Piromalis, K. G. Arvanitis, T. Bartzanas, D. Loukatos, Applications of IoT for optimized greenhouse environment and resources management, *Comput. Electron. Agric.*, **198** (2022), 1–32. <https://doi.org/10.1016/j.compag.2022.106993>
6. N. N. K. Krisnawijaya, B. Tekinerdogan, C. Catal, R. van der Tol, Multi-criteria decision analysis approach for selecting feasible data analytics platforms for precision farming, *Comput. Electron. Agric.*, **209** (2023), 1–11. <https://doi.org/10.1016/j.compag.2023.107869>
7. K. E. A. Saputro, L. Hasim, Karlinasari, I. S. Beik, Evaluation of Sustainable Rural Tourism Development with an Integrated Approach Using MDS and ANP Methods: Case Study in Ciamis, West Java, Indonesia, *Sustainability*, **15** (2023), 1–18. <https://doi.org/10.3390/su15031835>
8. M. F. B. Alam, S. R. Tushar, S. M. Zaman, E. D. S. Gonzalez, A. M. Bari, C. L. Karmaker, Analysis of the drivers of Agriculture 4.0 implementation in the emerging economies: Implications towards sustainability and food security, *Green Technol. Sustainability*, **1** (2023), 1–14. <https://doi.org/10.1016/j.grets.2023.100021>
9. I. D. Lopez, J. F. Grass, A. Figueroa, J. C. Corrales, A proposal for a multi-domain data fusion strategy in a climate-smart agriculture context, *Int. Trans. Oper. Res.*, **30** (2023), 2049–2070. <https://doi.org/10.1111/itor.12899>
10. L. Li, J. Lin, Y. Ouyang, X. R. Luo, Evaluating the impact of big data analytics usage on the decision-making quality of organizations, *Technol. Forecast. Soc. Change.*, **175** (2022), 1–14. <https://doi.org/10.1016/j.techfore.2021.121355>

11. A. Rejeb, A. Abdollahi, K. Rejeb, H. Treiblmaier, Drones in agriculture: A review and bibliometric analysis, *Comput. Electron. Agric.*, **198** (2022), 107017. <https://doi.org/10.1016/j.compag.2022.107017>
12. Y. Sheng, Z. Ma, X. Wang, Y. Han, Ethanol organosolv lignin from different agricultural residues: Toward basic structural units and antioxidant activity, *Food Chem.*, **376** (2022), 1–14. <https://doi.org/10.1016/j.foodchem.2021.131895>
13. M. Paul, M. Negahban-Azar, A. Shirmohammadi, H. Montas, Assessment of agricultural land suitability for irrigation with reclaimed water using geospatial multi-criteria decision analysis, *Agric. Water Manage.*, **231** (2020), 1–14. <https://doi.org/10.1016/j.agwat.2019.105987>
14. X. Cao, W. Zeng, M. Wu, X. Guo, W. Wang, The hybrid analytical framework for regional agricultural water resource utilization and efficiency evaluation, *Agric. Water Manage.*, **231** (2020), 1–14.
15. N. Chergui, M. T. Kechadi, M. McDonnell, The impact of data analytics in digital agriculture: a review, *O. Knowl. Adv. Technol.*, (2020), 1–13. <https://doi.org/10.1109/OCTA49274.2020.9151851>
16. M. Kavurmacı, C. B. Karakuş, Evaluation of irrigation water quality by data envelopment analysis and analytic hierarchy process-based water quality indices: The case of Aksaray City, Turkey, *Water Air Soil Pollut.*, **231** (2020), 1–17. <https://doi.org/10.1007/s11270-020-4427-z>
17. K. A. Shastry, H. A. Sanjay, Data analysis and prediction using big data analytics in agriculture, *Internet Things Anal. Agric.*, **2** (2020), 201–224. [https://doi.org/10.1007/978-981-15-0663-5\\_10](https://doi.org/10.1007/978-981-15-0663-5_10)
18. Z. Wang, J. Wang, G. Zhang, Z. Wang, Evaluation of agricultural extension service for sustainable rural development using a hybrid entropy and TOPSIS method, *Sustainability*, **13** (2021), 347–357. <https://doi.org/10.3390/su13010347>
19. X. Wang, Analysis and evaluation research on the influencing factors of the development of local agricultural products, in *E3S Web of Conferences*, **251** (2021), 1–4. <https://doi.org/10.1051/e3sconf/202125102096>
20. S. Talukdar, M. W. Naikoo, J. Mallick, B. Praveen, P. Sharma, A. Rahman, Coupling geographic information system integrated fuzzy logic-analytical hierarchy process with global and machine learning-based sensitivity analysis for agricultural suitability mapping, *Agric. Syst.*, **196** (2022) 1–13. <https://doi.org/10.1016/j.agry.2021.103343>
21. M. Raj, S. Gupta, V. Chamola, A. Elhence, M. Atiquzzaman, D. A. Niyato, Survey on the role of the Internet of Things in adopting and promoting Agriculture 4.0, *J. Network Comput. Appl.*, **187** (2021), 1–12. <https://doi.org/10.1016/j.jnca.2021.103107>
22. R. Sharma, S. S. Kamble, A. Gunasekaran, Big GIS analytics framework for agriculture supply chains: A literature review identifying the current trends and future perspectives, *Comput. Electron. Agric.*, **155** (2018), 103–120. <https://doi.org/10.1016/j.compag.2018.10.001>
23. O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, M. N. Hindia, An overview of the Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges, *IEEE Internet Things J.*, **5** (2018), 3758–3773. [10.1109/JIOT.2018.2844296](https://doi.org/10.1109/JIOT.2018.2844296)
24. C. Arora, A. Kamat, S. Shanker, A. Barve, Integrating agriculture and industry 4.0 under “agri-food 4.0” to analyze suitable technologies to overcome agronomical barriers, *Br. Food J.*, **124** (2022), 2061–2095. <https://doi.org/10.1108/BFJ-08-2021-0934>
25. A. Kumar, S. Pant, Analytical hierarchy process for sustainable agriculture: An overview, *MethodsX*, **10** (2023), 1–12. <https://doi.org/10.1016/j.mex.2022.101954>

26. T. H. Nguyen, D. Nong, K. Paustian, Surrogate-based multi-objective optimization of management options for agricultural landscapes using artificial neural networks, *Ecol. Modell.*, **400** (2019), 1–13. <https://doi.org/10.1016/j.ecolmodel.2019.02.018>
27. G. Latif, S. E. Abdelhamid, R. E. Mallouhy, J. Alghazo, Z. A. Kazimi, Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model, *Plants*, **11** (2022), 1–17. <https://doi.org/10.3390/plants11172230>
28. Y. Kittichotsatsawat, V. Jangkrajarn, K. Y. Tippayawong, Enhancing the coffee supply chain towards sustainable growth with big data and modern agricultural technologies, *Sustainability*, **13** (2021), 1–12. <https://doi.org/10.3390/su13084593>
29. L. N. Yang, Z. C. Pan, W. Zhu, E. J. Wu, D. C. He, X. Yuan, et al., Enhanced agricultural sustainability through within-species diversification, *Nat. Sustainability*, **2** (2019), 46–52. <https://doi.org/10.1038/s41893-018-0201-2>
30. K. F. Davis, J. A. Gephart, K. A. Emery, A. M. Leach, J. N. Galloway, P. Dorico, Meeting future food demand with current agricultural resources, *Glob. Environ. Change*, **39** (2016), 125–132. <https://doi.org/10.1016/j.gloenvcha.2016.05.004>
31. C. Maraveas, Incorporating artificial intelligence technology in smart greenhouses: Current state of the art, *Appl. Sci.*, **13** (2022), 1–35. <https://doi.org/10.3390/app13010014>
32. C. Maraveas, P. G. Asteris, K. G. Arvanitis, T. Bartzanas, D. Loukatos, Application of bio and nature-inspired algorithms in agricultural engineering, *Arch Comput. Methods Eng.*, **30** (2022), 1979–2012. <https://doi.org/10.1007/s11831-022-09857-x>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)