International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

# Henry Gas Solubility Optimization with Deep Learning Model for Crop Yield Prediction and Crop Type Classification

**P. S. S. Gopi[1], Dr. M. Karthikeyan[2]**

[1]Assistant Professor / Programmer, Department of Computer and Information Science, Faculty of Science, Annamalai University, Tamilnadu, India
*gopipss[at]gmail.com*

[2]Assistant Professor / Programmer, Department of Computer and Information Science, Faculty of Science, Annamalai University, Tamilnadu, India
*karthiaucse[at]gmail.com*

**Abstract:** ***Agriculture is considered to be the backbone of the Indian economy, with more than half of the country's population depending on agriculture. Crop production can be forecasted by utilizing machine learning (ML) methods rely upon parameters like meteorological conditions, rainfall, and crop. The powerful and most popular supervised ML technique, Random Forest, can do both regression and classification tasks. It can be used in crop selection for reducing crop yield output losses, irrespective of the distracting atmosphere. Meteorological conditions and other related environmental components bring a significant danger to the long-term viability of agriculture. ML is significant as renders a decision-support tool for Crop Yield Prediction (CYP), which will help to make decisions like which crops have to be cultivated and during the crop's growing season. This manuscript develops a new Henry Gas Solubility Optimization with Deep Learning Model (HGSO-DLM) technique to predict crop yield and classify crop types. In the presented HGSO-DLM model, two major processes are involved. At the initial level, the presented HGSO-DLM model employs a deep stacking auto-encoding (DSAE) model for yield prediction and crop classification. Next, in the second stage, the HGSO algorithm is applied for effectual hyper parameter optimization of the DSAE model. To exhibit the improvements of the HGSO-DLM model, a wide range of simulation results were performed and the comparison study reported the improvements of the HGSO-DLM model.***

**Keywords:** Henry gas solubility optimization; Deep learning; Crop type classification; Crop yield prediction; Metaheuristics

## 1. Introduction

Agriculture is considered to be the main sector as it provides a substantial amount of food. Presently, many counties are seeming to be hungry till now because of the lack or scarcity of food with an increasing populace [1]. The consolidated influences of an increasing populace, natural climate variability, soil loss, and climate-changing demand methods for ensuring production and crop growth in an opportune and dependable manner [2]. It even needs to donate to mount farming food production sustainability. Such necessities denoted that land valuation, crop protection, and crop yield estimation are highly important to worldwide food manufacture [3]. Therefore, a precise crop yield forecast was obligatory to depend on by the nation's politicians to get suitable import and export assessments to enhance national food security. But, because of many multifaceted issues, the forecast of crop yield was made difficult. Essentially, the crop yield was reliant on many aspects, including genotype, sceneries, soil quality, pest infestations, accessibility and eminence of water, climatic circumstances, harvest preparation, and many more [4]. Crop yield procedures and techniques are fundamentally nonlinear and time-specific. These plans are too compound as the combination of a big variety of unified influences which can be affected and described by non-arbitration and outside features [5]. Before, agriculturalists hinge on their knowledge and reliable historic data for crop yield forecasts and relaxed on the important farming choices as per the estimation. Yet, in current years, the development of novel inventions, counting crop method simulation and ML has seemed to forecast yield more exactly, together with the capability for examining an enormous volume of data utilizing high-performance computing [6].

Lately, ML methods were implemented for crop yield prediction which includes artificial neural networks, multivariate regression, association rule mining, and decision tree [7]. A salient feature of ML techniques was that they indulgence the output (crop yield) as an implied function of input variables (environmental components and genes), which is a highly complex and non-linear function [8]. The authors hired a neural network (NN) with one hidden layer for predicting corn yield utilizing input data on weather, soil, and management. ML was a technology that delivers schemes with the capability to automatically improve and learn from experience by recurrently training [9]. It comprises a set of well-defined techniques that gather precise statistics and smear exact procedures to attain anticipated outcomes. ML approaches were implemented in the agriculture field for enhancing the quality and productivity of the crops full-grown [10]. The techniques in ML were employed to regulate a specific crop under which circumstances the finest harvest is produced.

This manuscript develops a new Henry Gas Solubility Optimization with Deep Learning Model (HGSO-DLM) technique to predict crop yield and classify crop types. In the presented HGSO-DLM model, two major processes are involved. At the initial level, the presented HGSO-DLM model employs a deep stacking auto-encoding (DSAE) model for yield prediction and crop classification. Next, in the second stage, the HGSO algorithm is applied for effectual hyper parameter optimization of the DSAE model. To exhibit the improvements of the HGSO-DLM model, a wide range of simulation results were performed.

## 2. Related Works

Kim et al. [11] relate distinct AI techniques for developing an optimum crop yield predictive method in the Midwestern US. With the experimental for examining the impacts of phenology utilizing 3 distinct periods, the authors chose the July–August (JA) dataset as an optimum month for predicting soybean and corn yields. 6 various AI techniques for crop yield prediction were tested during this study. Afterward, an entire and objective comparison was conducted betwixt the AI techniques. Especially for the DNN technique, the author implemented an optimized procedure for ensuring optimum configurations for the layered infrastructure, activation function, cost function, drop-out ratio, and optimizer. Sinwar et al. [12] recognize the past developments and evolving AI-based approaches from precision agriculture definitely to yield predictive and smart irrigation. The AI-based method offers appropriate data on crop harvest at a primary step and their connected smart irrigation management method was effective in judicious utilization of important resources like energy and water for cultivation.

Alreshidi [13] discovers the present IoT or AI technical implemented for SSA and then, recognizes IoT/AI technologies structure able of supporting the progress of the SSA platform. Along with contributing to the present body of data, this study review investigation, and progress in SSA and offers an IoT/AI infrastructure for establishing a Smart, Sustainable Agriculture environment as a solution. Sharma et al. [14] concentrations on the consumption of predicting computational intelligence system to estimate nitrogen status from wheat yield. The estimation is dependent upon the examination of crop images taken in an area at different lighting illumination. The wheat yield was primarily exposed to HSI color normalization, and then the optimized method utilizing GA and ANN-based predictive and crop accuracy status classifier. This ANN-based optimization method is considerably distinguished betwixt the wheat yields in another unwanted plant and weed but recognizes the crop harvest age as a categorical class. Abbaszadeh et al. [15] examine an infrastructure that utilizes the Bayesian Model Averaging (BMA) and a group of Copula purposes for integrating the outcomes of multiple DNNs comprising the 3DCNN and ConvLSTM and offer a probabilistic evaluation of soybean crop harvest.

## 3. The Proposed Model

In this manuscript, a new HGSO-DLM technique has been developed to predict crop yield and classify crop types. In the presented HGSO-DLM model, two major processes are involved. At the initial level, the presented HGSO-DLM model employed the DSAE model for yield prediction and crop classification. Next, in the second stage, the HGSO algorithm is applied for effectual hyperparameter optimization of the DSAE model.

### 3.1. Yield Prediction and Crop Type Classification

The presented HGSO-DLM model employed the DSAE model for yield prediction and crop classification. The AE is an FFNN that has more than one hidden unit [16]. It is a kind of unsupervised neural network, wherein the network efforts to match output to input vector wherever possible. Furthermore, it is utilized to produce a lower or higher dimension depiction of the input dataset. The usage of unsupervised learning of compressed data encoding makes neural networks extremely versatile. Additionally, this network is trained single layer successively that minimalized the computation resource required for designing an efficient mechanism. Once the hidden layer is of lesser dimension when compared to the input and output layers, the network is utilized for encoding the dataset. Multi-layered AE is sequentially trained, which allows for the gradual compression of data, generating what is named an SAE. The self-encoding models consist of output, input, and hidden layers as follows:

$$x_i = [x_{i1}, x_{i2}, \dots, x_{ij}]2^T \tag{1}$$

Whereas i indicates the $i\text{-}th$ flow table feature vectors, and j signifies all the flow table features. The vector encompasses $j\text{-}th$ features. The hidden layers encode and compress the input features of the flow table based on the following expression:

$$encoder = W_1 x_i + b_1 \tag{2}$$

Now W1 indicates the weight interconnecting the input and hidden layers, $xi$ indicates the input features of $i\text{-}th$ flow table, and $b_1$ represent the bias of hidden layers. Afterward, the encoding is accomplished and defined on the output outcome of the hidden layers, the output layers are decoded and recreated to generate an output of similar size as input layers:

$$decoder = f(W_2(encoder)_i + b_2) \tag{3}$$

Here, $f$ indicates the activation function, $W_2$ denotes the weight between the hidden and output layers, (encoder)i shows the stream table feature compressed through the hidden units, and $b_2$ indicates the bias of output layers. Lastly, the aim of training the self-encoding mechanism can be accomplished by minimalizing the loss function as follows:

$$loss = \sum_{i=1}^{n} \left(x_i - (decoder_i)\right)^2 \tag{4}$$

From the expression, $n$ indicates the amount of flow table features, $x_i$ denotes the input flow table features, and (decoder) denotes the flow table features $x_i$ via self-encoding. To accomplish feature extraction and dimensionality reduction while creating the algorithm, the study presents a deep stack auto-encoding method. The DSAE method can be made using stacked the input and hidden layers of the self-encoding model. All the self-encoding models generate a hidden unit. Fig. 1 illustrates the structure of DSAE.

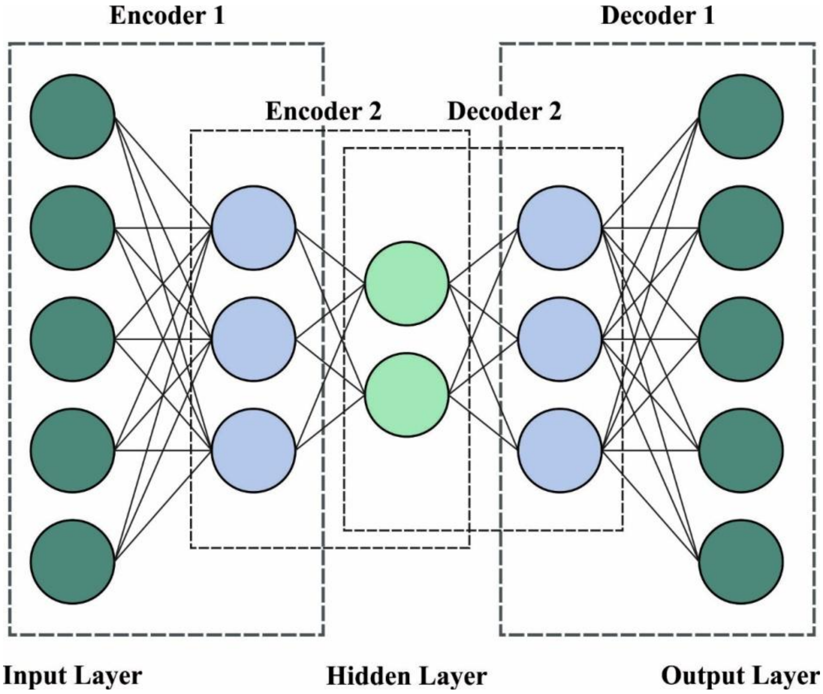


**Figure 1:** Architecture of DSAE

### 3.2 Hyperparameter Tuning using HGSO Algorithm

Here, the HGSO algorithm is applied for effectual hyperparameter optimization of the DSAE model. The HGSO technique is a novel metaheuristic technique introduced in 2019 [17]. This metaheuristic approach is created based on physical or biological phenomena. The probabilistic factor in the evolutionary procedure allows escaping from the local optima that offer the advantage of good generalization, simple operation, and generalization. The well-known physics law, Henry's law, stimulates the HGSO approach that describes the solubility phenomenon of the gas in a liquid under specific pressure. The mathematical expression of the abovementioned Henry's law can be given in the following:

Step 1: Initialization. $X_i$ location of $i^{th}$ gas particles in the initial population can be expressed as follows:

$$X_i^0(t+1) = lb + rand(\mathrm{n}, dim) * (ub - lb) \qquad (5)$$

Where, $ub, lb,$ and $dim$ show the upper limit, the lower limit, and the problem dimension, correspondingly. The preliminary value of Henry's constant $H_j$ for $j\text{-}th$ clusters, the $P_{\mathrm{i,j}}$the partial pressure of $i\text{-}th$ gas in $j\text{-}th$ clusters, and $C_j$the constant value of $j\text{-}th$ clusters is equated by the following:

$$H_j^0(t) = l_1 * rand(m,1), P_{i,j}^0 = l_2 * rand(n,1), C_j^0 = l_3 * rand(m,1) \quad (6)$$

From the expression, $m$ indicates the gas cluster count. $l_1, l_2, l_3$ denotes the constant equivalent to 5e-03, 100, le-02, correspondingly.

Step 2: Clustering. The gas particle with $n$ population is disseminated with $m$ cluster as gas types. All the clusters havethe same amount of candidate particles with the similar $H_j$ Henry's coefficient and $C_j$ constant values. Every cluster gas has a constant value $H_j$ and $C_j$.

Step 3: Evaluation. In all $j\text{-}th$ clusters, the better candidate particle $\chi_{j,best}$ that attains the better fitness values in $j\text{-}th$ clusters is assessed for finding the global better gas-particle $\chi_{g,best}$ amongst $n$ population.

Step 4: Upgrade Henry's coefficient. In distinct iterations and clusters, Henry's coefficient is upgraded to Henry's law as follows:

$$H_j(t+1) = H_j(t) \times exp\left(-C_j \times \left(\frac{1}{T^t} - \frac{1}{T^\theta}\right)\right) \quad (7)$$

Where $T^\tau = \exp\left(\frac{-t}{iter}\right)$ changes in all the iterations, $T$ demonstrates the temperature, and $T^\theta$ indicates a constant number of 298.15.

Step 5: Update solubility. The solubility $S_{\mathrm{i},j}$ of $i\text{-}th$ gas particle in $j\text{-}th$ clusters are arithmetically formulated by:

$$S_{i,j}(t) = K \times H_j(t) \times P_{i,j}(t) \qquad (8)$$

Where $K$ indicates a constant value equivalent to one.

Step 6: Updating location. The following location of $i^{th}$ gas particles in $j^{th}$ clusters are upgraded by the following equation [18]:

$$X_{i,j}(t+1) = X_{i,j}(t) + f \times rand \times (\phi_{i,j} \times (x_{j,best} - x_{i,j}(t)) + flag \times rand \times \alpha \times (S_{i,j}(t) \times x_{best}(t) - X_{i,j}(t))_t (\phi_{i,j} = \beta \times exp\left(-\frac{F_{best}(t) + \varepsilon}{F_{i,j}(t) + \varepsilon}\right) \quad (9)$$

From the expression, $f$ indicates a flag index equivalent to $-1$ or 1, that is employed for changing the direction of the searching agent. $rand$ indicates arbitrary numbers within zero and one, and each $rand$ signifies a dissimilar arbitrary number. $\phi_{i,j}$ represent the capability of $i\text{-}th$ gas particles in the $j$-$th$ clusters. $\alpha$ represent the effect of another gas particle on $i^{th}$ gas candidates that are fixed as 1. $\beta$ and $\varepsilon$ are constant coefficients equivalent to 1 and 0.05, correspondingly.

Step 7: Attain the worst agent. $N_w$ the worst agent can be ordered and applied in the optimization technique to prevent local optimal which is shown below:

$$N_w = n \times (rand \times (c_2 - c_1) + c_1) \tag{10}$$

Now, $c_1$ and $c_2$ is continually equivalent to 0.1 and 0.2, correspondingly. Each $rand$ function in the model representsan arbitrary vector within [0,1].

Step 8: Upgrade the worst location. The location of the worst particles is upgraded using an arbitrary number as follows:

$$X_w = lb + rand \times (ub - lb) \tag{11}$$

Afterward the abovementioned process, the location $X_{i+1}$ of $(i+1)^{th}$ gas particles are initialized.

## 4. Experimental Validation

This section examines the performance of the HGSO-DLM technique. Table 1 and Fig. 2 show the crop classification results of the HGSO-DLM technique. The results indicated that the HGSO-DLM technique has reached higher $accu_y$ of 98.07% whereas the NC-SAE, SVM-kernel, SVM, SSAE-CNN, PCA-CNN, and DT models have obtained lower $accu_y$ of 94.64%, 91.73%, 89.49%, 90.85%, 88.62%, and 85.07%.

**Table 1:** Comparative analysis of HGSO-DLM system with existing algorithms

| Methods | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| HGSO-DLM | 98.07 | 98.14 | 98.22 | 98.11 |
| NC-SAE Model | 94.64 | 94.06 | 94.78 | 95.49 |
| SVM-Kernel Model | 91.73 | 91.13 | 92.42 | 93.80 |
| SVM Model | 89.49 | 88.17 | 88.70 | 88.46 |
| SSAE-CNN | 90.85 | 93.86 | 90.60 | 92.94 |
| PCA-CNN | 88.62 | 89.27 | 87.75 | 89.08 |
| DT Model | 85.07 | 84.73 | 85.91 | 85.39 |

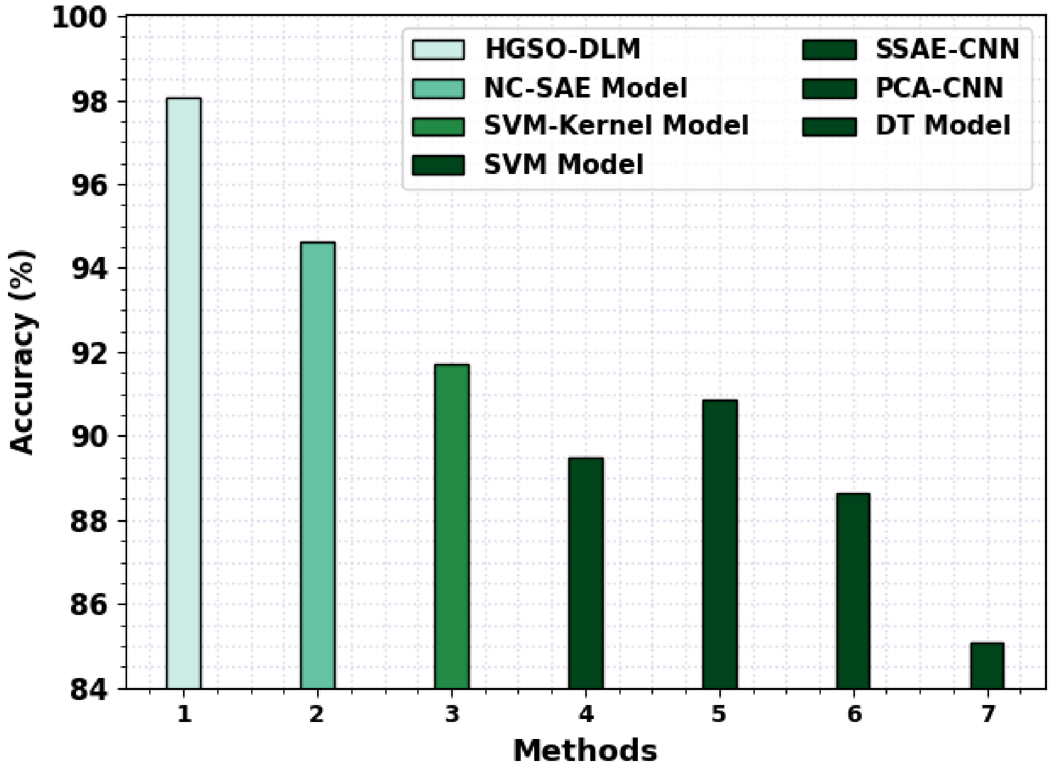


**Figure 2:** $Accu_y$ analysis of the HGSO-DLM system with existing algorithms

Fig. 3 reports a comparative crop classification performance of the HGSO-DLM technique. The experimental outcomes reported that the HGSO-DLM technique has surpassed other models. Based on $prec_n$, the HGSO-DLM technique has revealed increased $prec_n$ of 98.14% whereas the NC-SAE model has reached a reduced $prec_n$ of 94.06%. Meanwhile, based on $reca_l$, the HGSO-DLM method has exposed increased $reca_l$ of 98.22% whereas the NC-SAE approach has achieved reduced $reca_l$ of 94.78%. Eventually, based on $F1_{score}$, the HGSO-DLM methodology has exhibited increased $F1_{score}$ of 98.11% whereas the NC-SAE method has gained reduced $F1_{score}$ of 95.49%.



Paper ID: SR221014155237 DOI: 10.21275/SR221014155237

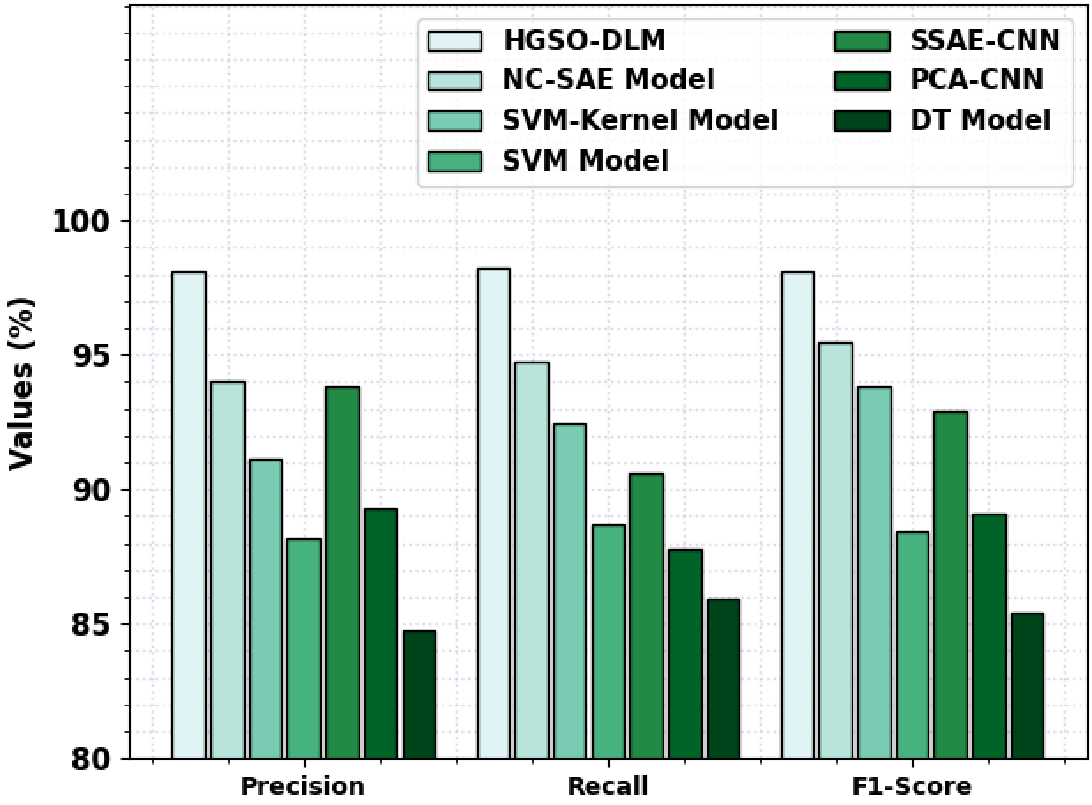


**Figure 3:** Comparative analysis of HGSO-DLM system with existing algorithms

The crop yield estimation results of the HGSO-DLM technique are examined in Table 2 and Fig. 4. The experimental values inferred that the HGSO-DLM technique has shown improved performance with an R2 score of 0.9961, RMSE of 0.8356, and MAE of 0.3011.

**Table 2:** Result analysis of HGSO-DLM system with distinct measures

| Metrics | Values |
|---|---|
| R2 Score | 0.9961 |
| RMSE | 0.8356 |
| MAE | 0.3011 |

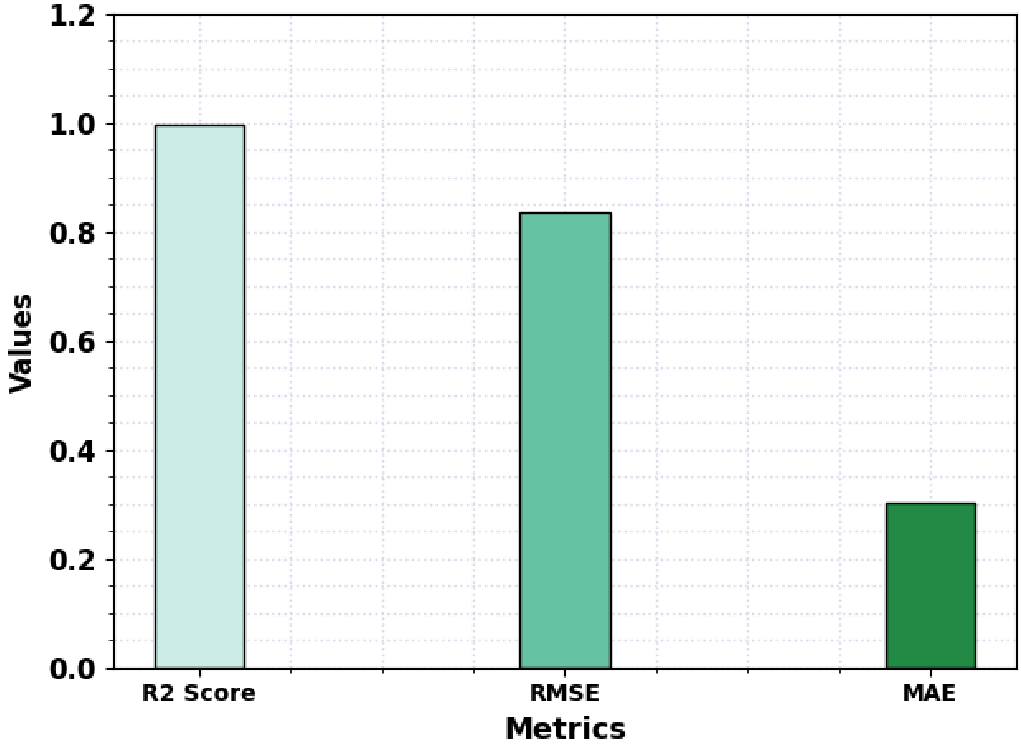


**Figure 4:** Result analysis of HGSO-DLM system with distinct measures

Table 3 and Fig. 5 demonstrate the comparative crop yield estimation results of the HGSO-DLM technique in terms of R2 score. The experimental values indicated that the KNN and MLR models have exhibited lower R2 score values of 87.05% and 89.10%. Along with that, the SVR and ANN models have accomplished reasonable R2 score values of 91.99% and 91.97% respectively. But the HGSO-DLM technique has shown maximum performance with an R2 score of 99.61%.

**Table 3:** R2 score analysis of HGSO-DLM system with other approaches

| Methods | R2 Score |
|---|---|
| HGSO-DLM | 99.61 |
| SVR Model | 91.99 |
| KNN Model | 87.05 |
| MLR Model | 89.10 |
| ANN Model | 91.97 |

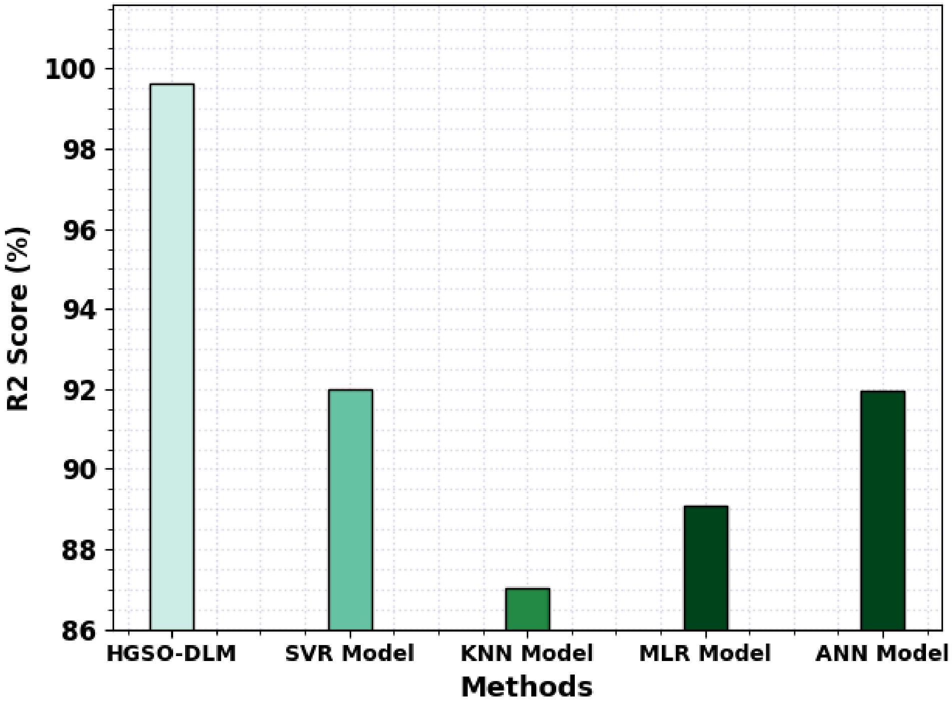


**Figure 5:** R2 score analysis of HGSO-DLM system with other approaches

## 5. Conclusion

In this manuscript, a new HGSO-DLM technique has been developed to predict crop yield and classify crop types. In the presented HGSO-DLM model, two major processes are involved. At the initial level, the presented HGSO-DLM model employed the DSAE model for yield prediction and crop classification. Next, in the second stage, the HGSO algorithm is applied for effectual hyperparameter optimization of the DSAE model. To exhibit the improvements of the HGSO-DLM model, a wide range of simulation results were performed and the comparison study reported the improvements of the HGSO-DLM model.

## References

[1] Ali, I., Cawkwell, F., Dwyer, E., Green, S., 2017. Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (7), 3254–3264

[2] Filippi, P., Jones, E.J., Wimalathunge, N.S., Somarathna, P.D.S.N., Pozza, L.E., Ugbaje, S.U., Bishop, T.F.A., 2019b. An approach to forecast grain crop yield using multilayered, multi-farm data sets and machine learning. Precis. Agric.

[3] Villanueva, M.B., Louella, M., Salenga, M., 2018. Bitter Melon Crop Yield Prediction using Machine Learning Algorithm. IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9.

[4] Saravi, B., Nejadhashemi, A.P., Tang, B., 2019. Quantitative model of irrigation effect on maize yield by deep neural network. Neural Comput. Appl. 1–14.

[5] Yalcin, H., 2019. An approximation for a relative crop yield estimate from field images using deep learning. In: 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). IEEE, pp. 1–6

[6] D. Elavarasan and P. M. D. Vincent, ''Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications,'' IEEE Access, vol. 8, pp. 86886–86901, 2020.

[7] Batool, D., Shahbaz, M., Shahzad Asif, H., Shaukat, K., Alam, T.M., Hameed, I.A., Ramzan, Z., Waheed, A., Aljuaid, H. and Luo, S., 2022. A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning. *Plants*, *11*(15), p.1925.

[8] Suresh, G., Kumar, A.S., Lekashri, S. and Manikandan, R., 2021. Efficient crop yield recommendation system using machine learning for digital farming. *International Journal of Modern Agriculture*, *10*(1), pp.906-914.

[9] Paudel, D., Boogaard, H., de Wit, A., van der Velde, M., Claverie, M., Nisini, L., Janssen, S., Osinga, S. and Athanasiadis, I.N., 2022. Machine learning for regional

crop yield forecasting in Europe. *Field Crops Research*, *276*, p.108377.

[10] Pham, H.T., Awange, J., Kuhn, M., Nguyen, B.V. and Bui, L.K., 2022. Enhancing Crop Yield Prediction Utilizing Machine Learning on Satellite-Based Vegetation Health Indices. *Sensors*, *22*(3), p.719.

[11] Kim, N., Ha, K.J., Park, N.W., Cho, J., Hong, S. and Lee, Y.W., 2019. A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern United States, 2006–2015. ISPRS International Journal of Geo-Information, 8(5), p.240.

[12] Sinwar, D., Dhaka, V.S., Sharma, M.K. and Rani, G., 2020. AI-based yield prediction and smart irrigation. In Internet of Things and Analytics for Agriculture, Volume 2 (pp. 155-180). Springer, Singapore.

[13] Alreshidi, E., 2019. Smart sustainable agriculture (SSA) solution underpinned by internet of things (IoT) and artificial intelligence (AI). arXiv preprint arXiv:1906.03106.

[14] Sharma, A., Georgi, M., Tregubenko, M., Tselykh, A. and Tselykh, A., 2022. Enabling smart agriculture by implementing artificial intelligence and embedded sensing. Computers & Industrial Engineering, 165, p.107936.

[15] Abbaszadeh, P., Gavahi, K., Alipour, A., Deb, P. and Moradkhani, H., 2022. Bayesian multi-modeling of deep neural nets for probabilistic crop yield prediction. *Agricultural and Forest Meteorology*, *314*, p.108773.

[16] Yu, J. and Yan, X., 2021. A new deep model based on the stacked autoencoder with intensified iterative learning style for industrial fault detection. *Process Safety and Environmental Protection*, *153,* pp.47-59.

[17] Hashim, F.A., Houssein, E.H., Mabrouk, M.S., Al-Atabany, W. and Mirjalili, S., 2019. Henry gas solubility optimization: A novel physics-based algorithm. *Future Generation Computer Systems*, *101*, pp.646-667.

[18] Ravikumar, S. and Kavitha, D., 2021. CNN-OHGS: CNN-oppositional-based Henry gas solubility optimization model for autonomous vehicle control system. *Journal of Field Robotics*, *38*(7), pp.967-979.