IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 3, June, 2026, pp. 1243-1256
Journal homepage: http://iieta.org/journals/ts

Multimodal Deep Learning for Tapioca Yield Prediction Using Field-Collected Sensor and Imagery Data

M. Arul Sankar[1*], K. Venkatasalam[2]

[1] Department of Information Technology, Mahendra Engineering College, Mallasamudram, Namakkal 637503, India
[2] Department of Computer Science and Engineering, Mahendra Engineering College, Mallasamudram, Namakkal 637503, India

Corresponding Author Email: venkatasalamk@mahendra.info

https://doi.org/10.18280/ts.430313

Received: 22 March 2026
Revised: 4 June 2026
Accepted: 15 June 2026
Available online: 30 June 2026

Keywords:
tapioca yield prediction, environmental, Unmanned Aerial Vehicle/satellite imagery, sensors, regression, deep learning

ABSTRACT

Tapioca is a vital root crop whose yield is influenced by multiple environmental, soil, and physiological factors. Traditional yield estimation methods depend upon statistical models and manual sampling, which often lack real-time accuracy and are time-consuming and labour-intensive. The combination of Internet of Things (IoT) sensor networks and Unmanned Aerial Vehicle (UAV)/satellite imagery, with deep learning (DL) models, gives a promising solution for accurate real-time yield prediction. In this study, sensor data (soil moisture, weather parameters, pH, NPK levels, temperature, and leaf chlorophyll content) are collected using devices like Davis Vantage Pro2, Decagon 5TE and SPAD-502, while image data are captured using DJI Phantom 4 Multispectral UAVs and Sentinel-2 satellite imagery. Pre-processing of sensor data involves missing value imputation, normalization, and feature selection using the Hybrid Frilled Lizard Osprey (HFLO) algorithm, while image data undergo resizing, augmentation, and filtering approach. Image based features are extracted by a position-attention DenseNet-201 model. Also, the features from image data and sensor data are fused using a concatenation mechanism. Finally, fully connected layers and improved support vector regression (ISVR) are used to identify the tapioca yield prediction. The integrated DL methods give higher accuracy than individual modalities for both modalities. The image-based model captures spatial and phenotypic variations, while the sensor-based model captures fine-grained environmental effects. The proposed approach obtained the MAE value of 0.0566, RMSE value of 0.654, and R^2 value of 99.5, demonstrating the potential of multimodal real-time data integration for precision agriculture in tapioca fields.

1. INTRODUCTION

Accurate crop production forecasts are crucial to precision agriculture because they enable effective decision-making in resource allocation, supply chain planning, and farm management [1, 2]. Tapioca yield prediction is particularly challenging due to the complex interactions between environmental, soil, and climatic characteristics, as well as variations in management strategies and data accessibility [3-5]. These challenges are made worse by the heterogeneous nature of agricultural data, which includes temporal sensor readings and geographical imaging data [6, 7].

Most conventional yield estimating methods rely on human observations and statistical models, which are frequently labor-intensive, time-consuming, and not real-time adaptive [8-10]. While deep learning (DL) and machine learning (ML) techniques have greatly increased prediction accuracy, they are still unable to capture complicated spatial-temporal correlations and efficiently integrate multimodal data sources [11-13]. While DL models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks require large datasets and involve high computational complexity [14, 15], ML models like Random Forest and Support Vector Machines (SVMs) have shown consistent performance but are limited in handling high-dimensional heterogeneous data [16, 17].

A significant technical challenge in tapioca production prediction is efficiently merging multimodal data that captures nonlinear environmental interactions and ensuring real-time prediction with good accuracy and generalization [18]. Because redundant and superfluous features in high-dimensional datasets can impair model performance, effective feature selection is essential [19]. To solve these problems, this study proposes a multimodal architecture that integrates Internet of Things (IoT) sensor data, aerial photography, and sophisticated feature selection and learning techniques. The suggested method combines an enhanced Support Vector Regression (SVR) model for prediction, attention-enabled DenseNet-201 for feature extraction, and Hybrid Frilled Lizard Osprey (HFLO)-based feature selection. While preserving computing efficiency, this methodology is intended to increase feature significance, capture spatial-temporal dependencies, and boost prediction accuracy [20]. To address these issues, the combination of IoT sensor networks and

Unmanned Aerial Vehicle (UAV)/satellite imagery, with DL models, gives a promising solution for accurate real-time yield prediction.

1.1 Motivation

Farming is most significant for enhancing food production and growth of a country's income world-wide. Several studies performed crop, rice, wheat, and paddy yield prediction, but were limited to tapioca yield prediction, which is much needed for agricultural growth. The existing studies obtained better results in agricultural systems; however, they faced some challenges like a lack of real-time accuracy, time-consuming, labour-intensive, and generalization. This motivated to develop DL model for real-time yield prediction through the collection of sensor data and satellite imagery. Thus, the proposed research developed novel fully connected layers and improved support vector regression (ISVR) model for tapioca yield prediction, thereby reducing the time- and labour-intensive challenges.

1.2 Contributions

By combining multimodal data with cutting-edge learning methods, this study seeks to provide a reliable and accurate framework for predicting tapioca yield. The following are this work's main contributions,

•To make use of multimodal data integration, which combines satellite imaging and IoT sensor data to provide more accurate and timely yield prediction than single-source methods.

•To increase data quality and model performance by designing an efficient data pre-processing pipeline that includes image enhancement techniques and sensor data normalization.

•To suggest a novel HFLO optimization technique for feature selection. In contrast to traditional techniques, HFLO combines the exploitation power of Osprey optimization with the exploration capacity of Frilled Lizard optimization (FLO), leading to improved feature subset selection and decreased redundancy.

•To identify the most interesting areas in imaging data and capture long-range spatial correlations using a Position Attention-based DenseNet-201 model for feature extraction improves the network's capacity.

•To create a yield prediction model using ISVR. The HFLO algorithm's refined hyper parameter tuning, improved kernel adaptation for managing nonlinear agricultural data, and better feature selection and parameter optimization all contribute to the improvement.

In this research organization section, as follows: 2 denotes the recent studies based on related work, 3 describes the proposed framework details, 4 illustrates the results & discussion, and 5 represents the conclusion & future scope.

2. RELATED WORK

Yield prediction based on some recent studies is as follows.

2.1 Methods based on machine learning

Traditional ML approaches have been frequently employed for crop production prediction because of their ease of use and efficiency. For example, a stacked ensemble model including Random Forest, AdaBoost, K-Nearest Neighbors, XGBoost, Elastic Net, and Linear Regression was presented by Ramesh et al. [21]. Nikhil et al. [22] compared Decision Trees, Random Forests, and Gradient Boosting models for agriculture productivity prediction in South India in an effort to increase forecast accuracy. The significance of feature engineering, which incorporates soil properties, climatic data, and remote sensing indices, was highlighted by their findings.

2.2 Crop-specific experimental research

Instead of using direct prediction modelling, other research concentrates on crop-specific experimental analysis. Cassava (tapioca) accessions in Tamil Nadu were assessed for yield, starch content, and resistance to Cassava Mosaic Disease by Kavitha et al. [23]. The study found high-performing cultivars and showed considerable genetic diversity. Such phenotypic and agronomic datasets are useful inputs for developing strong ML models for tapioca yield prediction, even though they are not predictive models.

2.3 Smart agriculture systems and Internet of Things

Smart farming solutions for yield monitoring and estimation have been made possible by recent developments in IoT. An IoT-based tuber harvesting robot with sensors, an ESP32 camera, and Arduino-based control was created by Prasanna et al. [24]. The method achieves high accuracy levels for yam and cassava crops by integrating predictive models that use decision trees and polynomial regression to forecast yield. This demonstrates how real-time sensing and automation are becoming increasingly important in agricultural forecast systems.

2.4 Deep learning and multimodal methods

More recently, multimodal DL approaches have shown superior performance by integrating heterogeneous data sources. Yewle et al. [25] introduced RicEns-Net, a deep ensemble model that combines synthetic aperture radar (SAR), optical imagery, and meteorological data for rice yield prediction. Their model effectively reduced feature dimensionality and achieved lower prediction error compared to state-of-the-art methods, demonstrating the benefits of multimodal data fusion. In a comprehensive and systematic evaluation of ML techniques for crop production prediction, Shawon et al. [26] emphasized the growing tendency toward DL and hybrid models for handling complicated agricultural information. Kalmani et al. [27] proposed a CNN-LSTM model that was enhanced with skip connections and attention mechanisms to capture both spatial and temporal dependency in yield prediction tasks. Menon et al. [28] developed a DL-based system for farm-level yield estimation using multi-temporal satellite pictures, showing that temporal feature extraction significantly enhances prediction performance in real-world scenarios. Similarly, Jiang et al. [29] demonstrated the efficacy of heterogeneous data fusion in precision agriculture by introducing a multimodal DL system that combines UAV imagery with meteorological data for cotton production prediction, attaining superior accuracy. Table 1 illustrates the summarization of baseline model performance and limitations.

Table 1. Summarization of baseline model performance and limitations

Authors and Ref.	Models	Performance	Limitations
Ramesh et al. [21]	(LR), Elastic Net, XGBoost Regressor, KNR, AdaBoost Regressor, and RFR	MAE - 7.20 tons/hectare	Scalability, time consumption, and generalization issues.
Kavitha et al. [23]	Machine learning (ML)	-	Labour-intensive and generalization challenges.
Prasanna et al. [24]	Polynomial regression and decision trees	R^2score-0.92	Time consumption and complexity issues.
Nikhil et al. [22]	Different ML models	-	Generalization and labour-intensive tasks.
Yewle et al. [25]	RicEns-Net	MAE- 336 kg/Ha	Complexity and scalability problems.
Shawon et al. [26]	ML and Deep learning (DL)	MAE- 7.20 tons/Hectare	No particular model implementation; no experimental validation.
Kalmani et al. [27]	CNN-LSTM with Skip and Attention Connections	R^2score-0.92	High computational complexity; a lot of training data is needed.
Menon et al. [28]	Using Multi-temporal Satellite Data for DL	RMSE- 0.78 tons/Hectare	Reliant on high-quality satellite data, restricted real-time applicability.
Jiang et al. [29]	Multimodal DL	R^2score-0.94	Complex architecture, higher computational expenses, and reliance on data.

Problem statement

Several studies performed crop, rice, wheat, and paddy yield prediction, but limited to tapioca yield prediction, which is much needed for agricultural growth. ML & DL models effectively performed for various yield prediction, however faced some challenges such as time consumption, labour-intensive, scalability, and generalization. In order to develop a DL model for real-time yield prediction through the collection of sensor data and satellite imagery. Thus, the proposed research developed novel fully connected layers and an ISVR model for tapioca yield prediction, thereby reducing the time and labour-intensive challenges.

3. PROPOSED METHODOLOGY

Tapioca is a vital root crop whose yield is influenced by multiple environmental, soil, and physiological factors. Traditional yield estimation methods depend upon statistical models and manual sampling, which often lack real-time accuracy and are time-consuming and labour-intensive. The combination of IoT sensor networks and UAV/satellite imagery, with DL models, gives a promising solution for accurate real-time yield prediction. In this study, sensor data (soil moisture, weather parameters, pH, NPK levels, temperature, and leaf chlorophyll content) are collected using devices like Davis Vantage Pro2, Decagon 5TE and SPAD-502, while image data are captured using DJI Phantom 4 Multispectral UAVs and Sentinel-2 satellite imagery. Pre-processing of sensor data involves missing value imputation, normalization, and feature selection using the HFLO algorithm, while image data undergo resizing, augmentation, Enhanced Median Filtering approach. Image based features are extracted by a Position Attention-aided DenseNet-201 model. Furthermore, the features from image data and sensor data using a concatenation mechanism. Finally, fully connected layers and ISVR are used to identify the tapioca crop yield prediction. The framework of the proposed methodology is denoted as Figure 1.

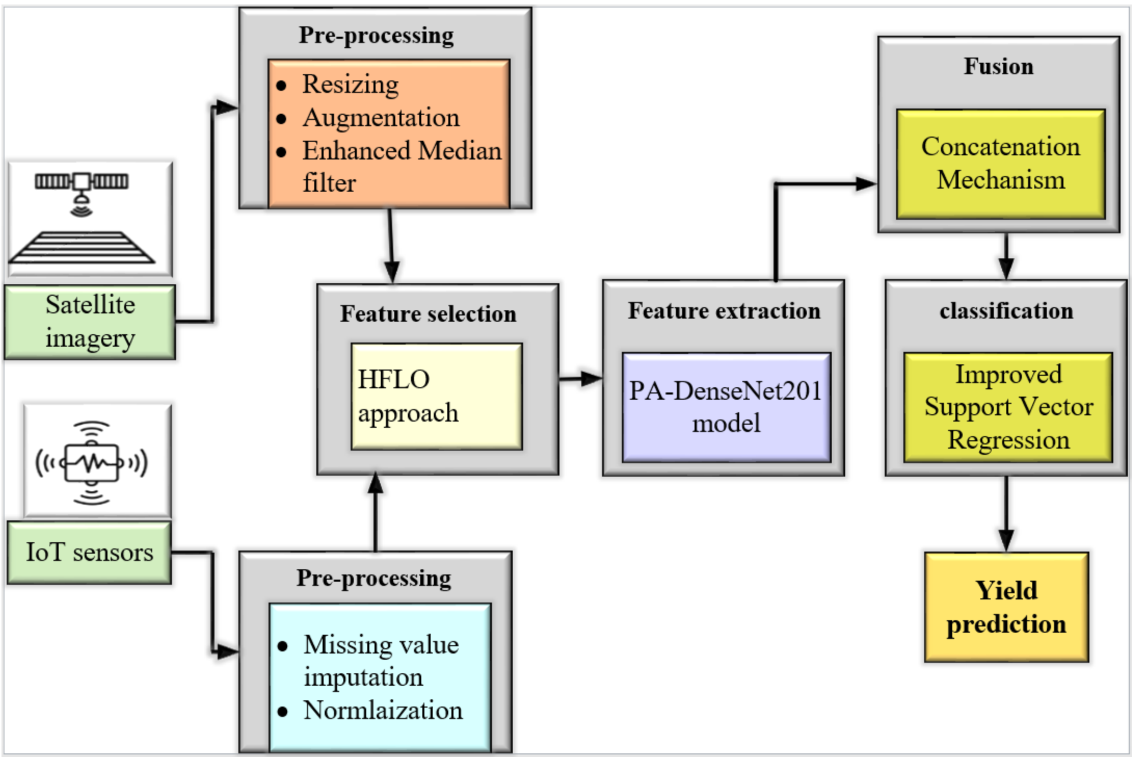


Figure 1. Framework of proposed methodology

3.1 Pre-processing

In this context, the preprocessing method is described to process the IoT sensor data and imagery data. The sensor data is efficiently preprocessed using missing value imputation and normalization methods. Here, image preprocessing for improving the quality and remove unwanted noise. Here, resizing, augmentation, and enhanced median filtering efficiently unify the strengths of preprocessing satellite images, resulting in enhanced image quality and better feature protection. This pre-processing step is essential for subsequent analyses in satellite data and IoT applications.

3.1.1 Sensor data processing

Missing value imputation: Due to hardware failures, transmission faults, or external interference, sensor data frequently contains missing or partial values. The integrity of downstream data analysis depends on how these missing values are handled. To confirm and close these gaps, imputation techniques are used [30]. The type of data and the anticipated effect on analysis accuracy determine the method to use.

Normalization: Normalization is a crucial stage in the preparation of sensor data, especially when the data comes from multiple sources of different sizes. This process transforms the data into a common scale without distorting variances in the range of values. Common normalization methods include min-max normalization, which scales data to a fixed range (typically 0 to 1), and z-score normalization, which ensures that all sensor features contribute equally to analysis, such as ML models [31], preventing any one variable from disproportionately influencing the results. Eq. (1) is used in min-max normalization to scale data to a specified range (0 to 1).

$$y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \tag{1}$$

Increases the stability of model training by maintaining the data distribution while guaranteeing that all features are on the same scale.

3.1.2 Imagery data processing

Resizing: Resizing is commonly used in sensor data processing, especially when dealing with time-series or image-based sensor outputs. It involves adjusting the dimensions or length of the data to a uniform size, which is essential for consistent input into analytical models or neural networks [32]. For instance, in time-series analysis, sensor readings may be resampled to a fixed time interval, ensuring that each sample has the same number of data points. In image data from sensors (e.g., thermal or optical sensors), resizing ensures compatibility with model input dimensions while maintaining the aspect ratio or key features of the original data.

Augmentation: Augmentation refers to the process of artificially increasing the diversity and volume of sensor data by applying various transformations. This is particularly useful in training ML models, where more varied data can improve generalization and robustness. In the context of sensor data, augmentation techniques may include adding noise, shifting time windows, rotating or flipping images, scaling values [33], or simulating environmental variations. These transformations help models learn to be resilient to real-world variability and reduce overfitting on limited datasets.

3.1.3 Enhanced median filtering

In contrast, utilizing enhanced median filtering is utilized to improve image quality and to reduce unwanted noise [34]. Although regular median filtering is also used for this purpose, it tends to produce more errors in this case. To solve that, this research uses enhanced median filtering. Next, bringing in the average error point as a metric helps reduce the error in the images and gives better results. The error point metrics as follows:

$$A = \frac{1}{M \cdot F} \sum_{x=1}^{M} \sum_{y=1}^{F} | o_{x,y} - \hat{o}_{x,y} | \tag{2}$$

where, $o_{x,y}$ is the image at trace x, point o in the original image, and $\hat{o}_{x,y}$ denotes the filter image. Here, the number of traces is represented as M, and N is the number of samples per trace. Furthermore, the outputs indicate that median filtering approaches perform the best, with the modified adaptive median filter having the minimum average error.

3.2 Feature selection using Hybrid Frilled Lizard Osprey optimization

In order to reduce redundancy and improve prediction accuracy, the suggested HFLO algorithm is used for optimal feature selection. The presented HFLO approach incorporates the exploration ability of the FLO and the exploitation robustness of the osprey optimization; the hybrid approach obtains a better balance between global search and local search. This helps in avoiding premature convergence and ensures a more accurate and diverse selection of optimal features. As a result, the approach improves predictive performance, reduces computational complexity, and enhances convergence speed compared to employing individual algorithms alone. Finally, this presented hybrid approach offers a strong and efficient solution for extracting the most relevant features from the collected database.

3.2.1 Initialization

Initialize the population of candidate feature subsets by simulating the behaviours of frilled lizard and ospreys optimization, aiming for a balance between exploration and exploitation. Define the factors: D is the total number of features in the database, population is P, $A \in \{0,1\}^{P \times D}$ is a binary population matrix where each row represents a feature subset.

3.2.2 Exploration using Frilled Lizard Optimization [35]

The HFLO algorithm blends the exploitation power of the Osprey Optimization Algorithm (OOA) with the exploration capacity of the FLO. While local refinement is carried out in the exploitation phase to enhance convergence toward optimal feature subsets, candidate solutions are updated using global search strategies during the exploration phase to prevent local optima. Each agent's position is updated iteratively as:

$$CI_i = \{P_F : O_F < O_i \text{ and } F \neq i\} \tag{3}$$

where, CI is the candidate feature set for the i^{th} FL, P_F denotes the population feature with a better O_k is the objective function value than the i^{th} FL, and $i = 1,2,\dots,N$ along with $k \in \{1,2,\dots,N\}$, F represents the features. The FLO algorithm assumes that each searcher randomly selects one feature from the set of selected feature positions and proceeds to predict. The movement of the searcher toward the candidate feature is mathematically modelled, and this approach is employed to update each individual's position in the population, as defined in Eq. (4). If the new position has a better objective value, it becomes the individual's new position, as shown in Eq. (5).

$$p_{i,d}^{x1} = p_{i,d} + r \cdot (SI_{i,d} - t \cdot p_{i,d}), \quad d = 1, 2, \ldots, m \tag{4}$$

$$P_i = \begin{cases} P_i^{x1}, & O_i^{x1} < O_i \\ P_1, & else \end{cases} \tag{5}$$

Let P_i^{x1} is the new suggested position of i^{th} searcher based on the first phase of FLO, $p_{i,d}^{x1}$ denotes its d^{th} dimension, O_i^{x1} is the objective function amount, r is a random number with a normal distribution from the interval [0, 1], O_i represents the i^{th} objective function, $SI_{i,d}$ is the d^{th} dimension of the selected feature for i^{th} searcher, t denotes the randomly taken from the set {1,2}, and m represents the number of decision variables.

3.2.3 Osprey using exploitation [36]

This phase deeply explores the features and discovers the features to catch. In a simulation of this real conduct, the second degree of updating the records in the OOA is modeled. The hunt agent's function in the search space is created by means of small adjustments resulting from modeling the identity of the function to the proper role. The OOA's exploitation improved power inside the neighborhood search and caused convergence closer to higher solutions towards the found solution. In the OOA design, a new random role was initially decided for each member of the data. This simulated the herbal conduct of search agents. The exploitation segment is given as:

$$a_{i,j}^{P2} = x_{i,j} + \frac{lb_j + r \cdot (ub_j - lb_j)}{t}, \; j = 1, 2, \ldots\ldots\ldots, m, \quad t = 1, 2, \ldots\ldots., T \tag{6}$$

$$A_{i,j}^{P2} = \begin{cases} a_{i,j}^{P2}, & lb_j \le a_{i,j}^{P2} \le ub_j; \\ lb_j, & a_{i,j}^{P2} < lb_j; \\ ub_j, & a_{i,j}^{P2} > ub_j, \end{cases} \tag{7}$$

$$A_i = \begin{cases} A_i^{P2}, F < F_i; \\ A_i, else, \end{cases} \tag{8}$$

where, A_i^{P2} is the i^{th} search agent's original place based on the next phase of the OOA, r represents the random number, lb represents the lower bound, ub represents the upper bound, $a_{i,j}^{P2}$ is the j^{th} aspect, F is its fitness role, $r_{i,j}$ are random numbers in the range [0,1], t is the iteration number of the method, and T is the total number of iterations. The fitness function is calculated as:

$$F = \lambda * E + (1 - \lambda) * \frac{|S|}{d} \tag{9}$$

where, E signifies the prediction error, $|S|$ signifies the number of selected features, and $\lambda \in [0,1]$ provides a balance between feature reduction and accuracy. The population size $N = 30$, maximum iterations $T = 100$, and control parameters $\alpha = 0.5$, $\beta = 0.3$ constitute the algorithm's set parameters. In order to balance convergence speed and solution quality, these parameters were chosen through empirical tweaking. For moderate feature dimensions, the HFLO algorithm's computing complexity of $O(N \times T \times d)$ is acceptable. All things considered, the HFLO algorithm offers a good trade-off between exploration and exploitation, resulting in the selection of an ideal subset of pertinent characteristics that enhances model generalization and lowers computing cost. Algorithm 1 provides the HFLO optimization pseudo code.

Algorithm 1. Pseudo code of Hybrid Frilled Lizard Osprey optimization

Input: dataset D, objective function $f(x)$, population size P, maximum iterations T

Step 1: Initialization

1. Initialize a population of P search members (Frilled lizard) $A = \{a_1, a_2, \ldots, a_P\}$.
2. Evaluate fitness $f(a_i)$ for each a_i.
3. Identify the best search member a_{best} with minimum $f(a)$.

Step 2: Exploration phase (FLO)

4. For each searcher $i = 1$ to P do
5. Construct the candidate feature set $CI_i = \{a_F : O_F < O_i \text{ and } F \neq i\}$.
6. Randomly select one feature a_f from C_i.
7. Update the position using predicting strategy:
$p_{i,d}^{x1} = p_{i,d} + r \cdot (SI_{i,d} - t \cdot p_{i,d})$, where $r \sim P(0,1)$.
8. If $f(a_i^{new}) < f(a_i)$ then
$a_i = a_i^{new}$.
9. End if
10. End for

Step 3: Exploitation phase (OOA)

11. For each searcher $i = 1$ to P do
12. Update position using OOA exploitation rule: Eq. (6)
13. If $f(a_i^{t+1}) < f(a_i^t)$ then
$a_i = a_i^{t+1}$
14. **End if**
15. **End for**

Step 4: Update and selection

16. Evaluate fitness for all agents using Eq. (9).
17. Update a_{best} if a better solution is found.

Step 5: Termination

18. Repeat Steps 2–4 until T iterations are completed.
19. Return the optimal feature subset $F^* = a_{best}$.

3.3 Position attention – DenseNet-201 model for feature extraction

Employing the suggested feature extraction model (Position Attention DenseNet-201) helps the model understand tapioca images better. DenseNet-201 model was better at extracting deep features from images while enhancing captured information. The position attention module (PAM) identifies long-range dependencies, enables efficient learning, and finds significant regions like diseased leaves, specific plant shapes, or textures. This effectiveness enables the model to focus on the tapioca areas and ignore the background (i.e., soil moisture, weather parameters, pH, NPK levels, temperature, and leaf chlorophyll content). Subsequently, the suggested model accurately predicts health conditions or growth stages. The integration enhances accuracy, efficient learning, and reduces computational power.

3.3.1 DenseNet-201

This model is an effective CNN widely employed for feature extraction in DL challenges, particularly in computer vision (Figure 2). Its framework is characterized by dense connections; here each layer receives inputs from all preceding layers, promoting feature reuse and efficient gradient flow. This framework enables DenseNet-201 to extra high [37], hierarchical features with fewer parameters compared to other deep networks like ResNet152 [38], making it both precise and relatively lightweight. When utilized as a feature extractor, the final classification layer is removed, and the output from the last convolutional block or global average pooling layer is utilized as a high-quality feature vector. These features are particularly effective in applications such as medical imagery, plant leaves, fine-grained classification, and object detection, often outperforming the effectiveness of deep models. Although integrating a position attention model in the DenseNet-201 model, it demands reduced complexity due to its dense connectivity, DenseNet-201 offers excellent generalization and has shown competitive performance on benchmark databases, making it a reliable choice for transfer learning and feature extraction pipelines.

3.3.2 Position attention

Image understanding depends significantly on discriminative feature representations, which can generally be improved by the inclusion of long-range contextual cues. To suggest adding a PAM. As shown in Figure 2, this module is tailored to learn spatial dependencies over the whole image, allowing the model to represent rich contextual relations beyond local domains. In doing so, the PAM actually boosts the representational capacity of the local features. In the next sections, explain how the PAM adaptively sums up spatial contextual information and achieves better image understanding task performance.

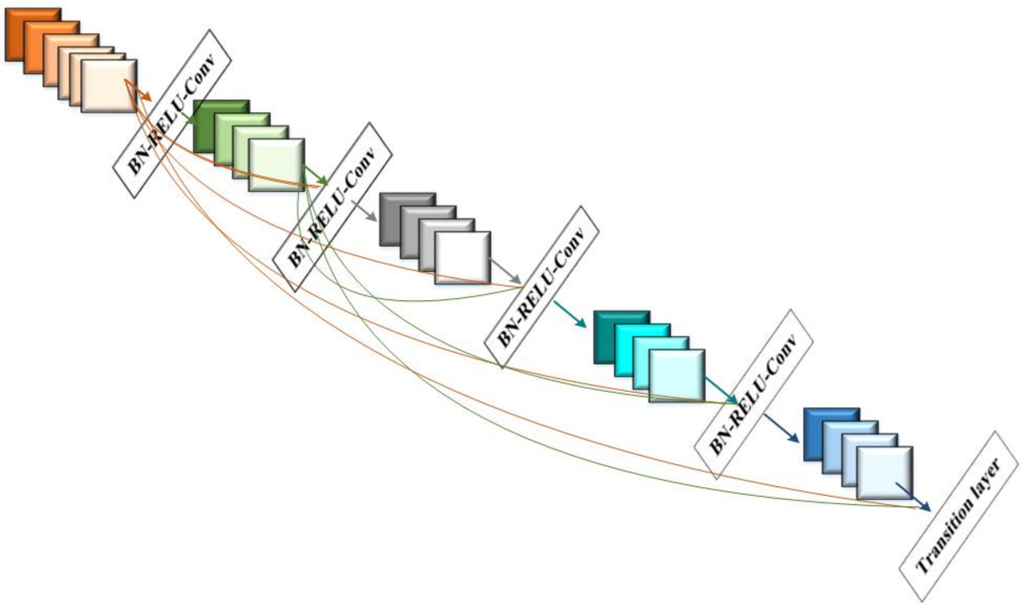


Figure 2. DenseNet-201 model

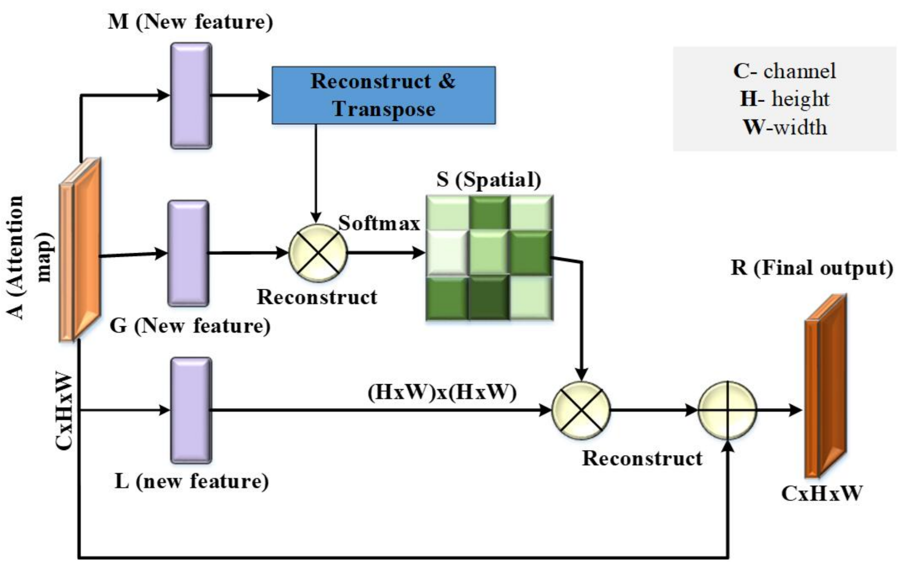


Figure 3. Position attention module (PAM)

Figure 3 shows the local feature $P \in \Re^{c\times h\times w}$ is fed into convolution layers to build two new feature maps G and M, separately, $\{G, M\} \in \Re^{c\times h\times w}$. Following the restructure them to $\Re^{c\times N}$, where $N = h \times w$ represents the number of pixels. Afterward, to calculate the attention map $A \in \Re^{N\times N}$, multiply the transpose of M and G use a softmax layer.

$$a_{yx} = \frac{\exp(G_x \cdot M_y)}{\sum_{x=1}^{N} \exp(G_x \cdot M_y)} \tag{10}$$

Here a_{jx} quantifies how the xth position affects the yth position. The two feature representations that are more alike at a position increase their correlation with one another. At the same time, transmit feature P into a convolution layer to build a new feature map $F \in \Re^{c\times h\times w}$ and reshape. Send it to $\Re^{c\times N}$. Next, multiply L by the transpose of A and reshape the result into $\Re^{c\times h\times w}$ matrix. The final output $R \in \Re^{c\times h\times w}$ is obtained by multiplying it by a scale parameter α and performing an element-wise sum operation on the features P.

$$R_y = \alpha \sum_{x=1}^{N} (a_{yx} L_x) + P_y \tag{11}$$

Here, α is initially set to 0 and subsequently learns to attach increasing weight. Eq. (11) indicates that the feature R at each place is a weighted sum. All the features in all positions and original features.

3.4 Concatenation mechanism for feature fusion

Concatenation is a fundamental feature fusion method in multisource DL employed to integrate heterogeneous data sources into a unified field representation. In contrast, sensor data and UAV multispectral imagery capture complementary information about tapioca and environmental conditions. The developed concatenation mechanism enables the model to exploit both spatial (image-based) and non-spatial (sensor-based) features jointly for enhanced prediction accuracy.

Mechanism: Let $I_s \in \Re^{n_s}$ is the IoT sensor data vector, and $I_x \in \Re^{h\times w\times c}$ is the UAV/satellite imagery patch with B spectral bands. Each modality is first passed through its own encoder network, sensor feature embedding as follows:

$$s_z = f_z(I_z) \in \Re^{d_z} \tag{12}$$

Here, f_z denotes the fully connected neural network (MLP) for tabular sensor data, s_z is the dense feature embedding of sensor data. Image embeddings are expressed in Eq. (13):

$$s_x = f_x(I_x) \in \Re^{d_x} \tag{13}$$

where, CNN or pretrained image encoder for extracting features from multispectral images is denoted as f_x, and s_x is the image feature embedding. Concatenation (fusion layer): The embeddings from both modalities are concatenated to form a joint feature vector:

$$s_f = [s_z \parallel s_x] \in \Re^{d_s + d_x} \tag{14}$$

Here, vector concatenation is represented as $\parallel$, s_f is the concatenated multimodal. The method vector captures both conditions, allowing the model to learn interactions. As a result, this method offers simplicity and effectiveness, while being well-suited for easy-to-implement and compatible neural network architectures. Features and network architecture dimensions are given in Table 2.

Table 2. Features and network architecture dimensions

Component	Layer Details	Output Dimension
Connected neural network (MLP)	Input Layer (Sensor Features)	d_s
	Dense Layer + ReLU	64
	Dense Layer + ReLU	128
	Output Embedding	128
	DenseNet-201 + Attention	256
	Global Average Pooling	256
	Output Embedding	256
Fusion layer	Concatenation	384
Fully connected	Dense Layer + ReLU	128
	Dense Layer	64
Prediction layer	Improved Support Vector Regression (SVR)	Output (Yield)

3.5 Tapioca yield prediction using improved support vector regression

In this study, an improved SVR model is utilized for tapioca yield prediction. To improve the learning ability, initially, integrates a CNN model [39] with SVR, because its performed efficient and reduces the time consumption. This approach is a supervised learning algorithm derived from SVMs [40, 41], developed for predicting continuous values rather than class labels. The main thing in SVR is to find a function that approximates the data within a certain ε margin of tolerance, while maintaining model simplicity. As a mathematical and statistical theory-based regression model, the objective of SVR is to create a functional relationship between input variables $a = (a_1, a_2, a_3, \dots, a_k)^T$ and the respective response b. Through the introduction of the non-linear feature mapping function $\psi(a)$, SVR is able to map the input data into high dimensional linear space. Thus, the relationship can be represented as a linear regression function in the high-dimensional space.

$$\hat{b}(a) = \omega^T \psi(a) + s \tag{15}$$

Here, $\hat{b}(a)$ is the approximation response, the weighted vector denoted as ω, and s is the bias term. To develop a well-generated strategy, the ε-insensitive constraint is employed. Given a training dataset or subset (a_x, b_x), where $a = 1,2,3, \dots, n$, the linear regression task can be reformulated as a constrained convex problem.

$$\min = \frac{1}{2}\|\omega\|^2; \begin{cases} \omega^T \psi(a^x) + s - b^x \le \varepsilon \\ b^x - \omega^T \psi(a^x) - s \le \varepsilon \\ x = 1, \dots, n \end{cases} \tag{16}$$

Eq. (16) is reformulated by incorporating the slack variables ξ_x^+ and ξ_x^- as follows:

$$\min = \frac{1}{2}\|\omega\|^2 \tag{17}$$

$$+C\sum_{x=1}^{n}\xi^{+(x)}+\xi^{-(x)};\begin{cases}\omega^T\psi(a^x)+c-b^x\le\varepsilon+\xi^{-(x)}\\ b^x-\omega^T\psi(a^x)-s\le\varepsilon+\xi^{-(x)}\\ \xi^{+(x)},\xi^{-(x)}\ge 0\\ x=1,\dots,n\end{cases}$$

Here, the regularized parameter is represented as $C > 0$. Utilizing Lagrangian theory, the dual formulation of Eq. (17) can be obtained.

$$\max=\begin{cases}-\frac{1}{2}\sum_{x,y=1}^{n}\left(\alpha^{+(x)}-\alpha^{-(x)}\right)\left(\alpha^{+(y)}-\alpha^{-(y)}\right)\\ k\left\langle a^x,a^y\right\rangle+\sum_{x=1}^{n}\left(\alpha^{+(x)}-\alpha^{-(x)}\right)b^x\\ -\sum_{x=1}^{m}\left(\alpha^{+(x)}+\alpha^{-(x)}\right)\varepsilon\end{cases} \tag{18}$$

$$\begin{cases}\sum_{x=1}^{n}\left(\alpha^{+(x)}-\alpha^{-(x)}\right)=0\\ 0\le\alpha^{+(x)},\alpha^{-(x)}\le C\\ x=1,\dots,n\end{cases}$$

Let the kernel function is $k\langle a^x, a^y\rangle$. Compute α_x^+ and α_x^- are initially obtained from Eq. (19).

$$\omega=\sum_{x=1}^{n}\left(\alpha_x-\alpha_x^*\right)*(a_x) \tag{19}$$

$$f(a)=\sum_{x=1}^{n}(\alpha_x-\alpha_x^*)k(a_x,a)+b \tag{20}$$

Here, $f(a)$ denotes the final regression output, $k(a,a') := < \phi(a)(a') >$ and ϕ represent the mapping function, k is the kernel function, and b is the respective response. Finally, fully connected layers to identify the tapioca crop yield prediction, while efficient results enhance performance.

The term improved SVR in this work refers to improving the standard SVR by fine-tuning parameters and optimizing feature representation. The fused multimodal feature vector created by merging sensor and image information is fed into the SVR model. In particular, satellite imagery is processed using an attention-based DenseNet-201 to generate a 256-dimensional feature vector after IoT sensor data is encoded into a 128-dimensional vector. These are combined to create a single feature vector that is fed into the SVR model. Table 3 uses an RBF kernel with hyper parameters that are optimized by the HFLO technique. Therefore, rather than changing the fundamental SVR formulation, the improvement is found in the use of deep multimodal feature fusion and better parameter selection, leading to improved prediction accuracy and generalization. Figure 4 depicts the ISVR model for yield prediction.

4. EXPERIMENTAL RESULTS OUTCOME

This research aims to propose "A Novel DL aided Regression Mechanism for Multi-Modal Tapioca Crop Yield Prediction" was proposed. The research database detail mention in section 4.1, and implemented by PYTHON tool. Then, the evaluation metrics are described in section 4.2, and the following section 4.3 denotes the analysis of tapioca yield prediction performance. Furthermore, a comparison model based on the discussion is given in section 4.4, and the conclusion & future scope are in section 5. System configuration is given in Table 3, and hyper parameter details are given in Table 4.

Table 3. System configuration

Parameter	Detail
Device name	SS094.smg.local
Processor	Intel ® Core ™ i7-4770 CPU @ 3.40 GHz
Installed RAM	16.0GB (15.9 GB usable)
System type	64-bit operating system, x64-based processor

Table 4. Hyper parameter details

Component	Hyper Parameter	Value
DenseNet-201	Learning rate	0.001
	Optimizer	Adam
	Epochs	50
	Batch size	32
Attention Module	Activation function	ReLU
Hybrid Frilled Lizard Osprey (HFLO) Algorithm	Population size (N)	30
	α (Exploitation)	0.5
	β (Exploration)	0.3
	Iterations (T)	100
Improved Support Vector Regression (SVR)	Regularization	1.0
	Kernel function	RBF
	Output layer	Fully connected layer

4.1 Dataset collection

For this study, a comprehensive multi-source dataset was collected to enable accurate and rapid tapioca yield forecasting. Over the course of a full crop growing season, data was gathered from an experimental agricultural field. A strong multimodal learning framework is made possible by the dataset's integration of IoT sensor measurements, UAV based photography, satellite data, and ground truth yield information. Devices like the Davis Vantage Pro2, Decagon 5TE, and SPAD-502 were used to gather the IoT sensor data. Important environmental and soil data, such as soil moisture (5–10%), soil temperature (10–40 ℃), soil pH (~4.5), NPK levels, air temperature (5–45 ℃), humidity, rainfall, sun radiation, and leaf chlorophyll content, were recorded by these sensors. About 500 sensor records were produced as a result of the data being recorded at various temporal resolutions, such as two-hour intervals and sporadic weekly readings.

Additionally, DJI Phantom 4 UAVs were used to get high resolution multispectral data in order to extract vegetation indices including NDVI, GNDVI, and NDRE. In addition, Sentinel-2 satellite imagery, which contributed about 6000 samples and allowed the extraction of indices like EVI and LSWI, offered wider spatial coverage with a resolution of 10-20 m. For the purpose of training and validating the model, ground truth yield data, expressed in kg/ha, were manually gathered at harvest. All data sources were georeferenced and synchronized chronologically to ensure consistency and effective integration. This large dataset provides sufficient geographical, temporal, and spectral variability, making it suitable for training DL models such as DenseNet-201 for accurate yield prediction.

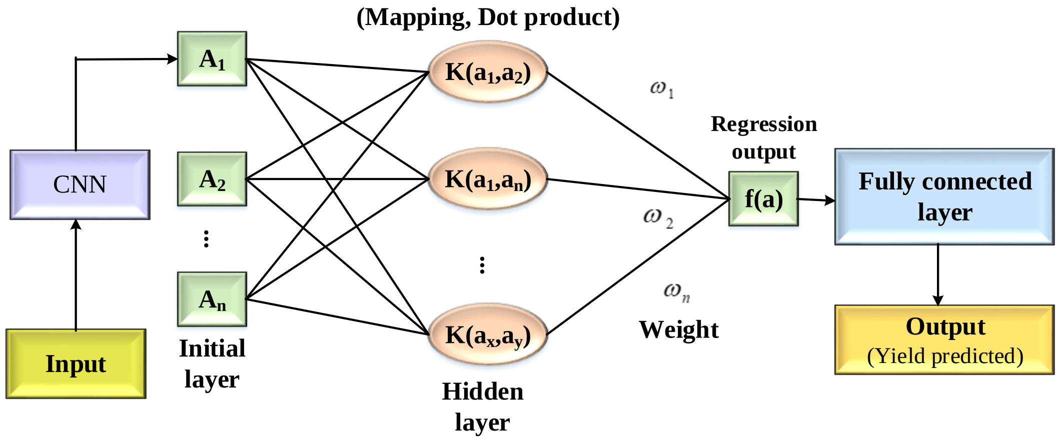


Figure 4. Improved Support Vector Regression (SVR) model for yield prediction

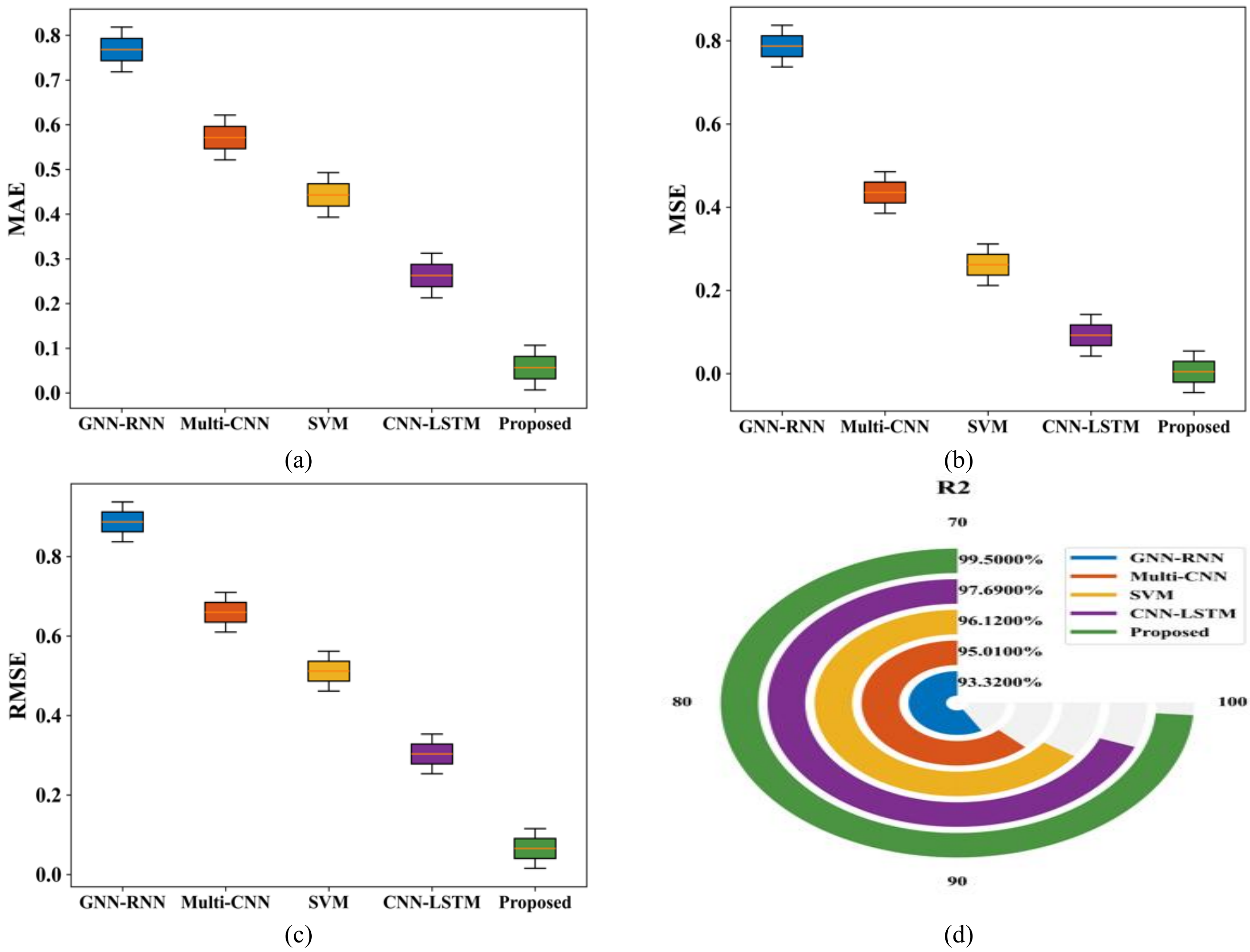


Figure 5. Yield prediction model performance comparison using assessment measures Mean Average Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R²

4.2 Evaluation metrics

The suggested method's efficacy in predicting tapioca production was assessed by a results analysis. Training and testing times, time consumption, and performance indicators such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Average Error (MAE), and R^2 [37] were used.

4.3 Analysis of tapioca yield prediction performance

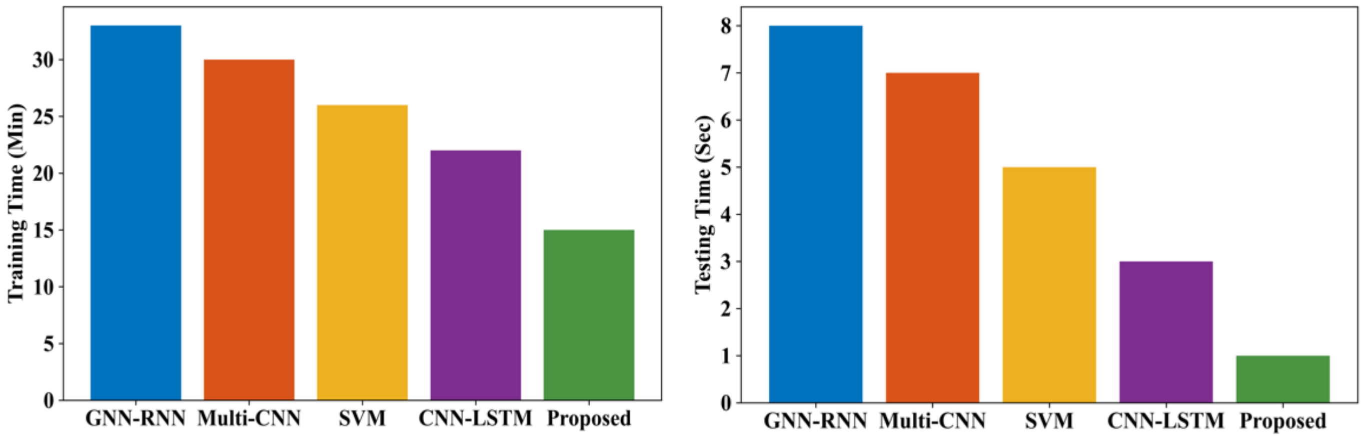
The four categories of metrics performance analysis of tapioca yield prediction are displayed in Figure 5.

Figure 5 displays the comparative performance analysis of various models for tapioca yield prediction utilizing a variety of evaluation measures, including MAE, MSE, RMSE, and R^2. By showing the standard deviation over several runs, error bars show each model's stability and variability. To guarantee a fair comparison, these results were obtained under similar experimental settings using the same training and testing dataset. The suggested Improved SVR model regularly produces the lowest error values and the highest R^2 score, demonstrating greater predictive accuracy, as seen in the figure. CNN-LSTM, on the other hand, performs comparatively well since it can handle temporal dependencies, although it has less generalization and a larger computational complexity. While Graph Neural Network-Recurrent Neural Network (GNN-RNN) and SVM perform moderately, their ability to effectively handle large-scale and multimodal data is limited. Although the Multi-CNN model efficiently captures spatial data, its higher computational burden makes it less appropriate for real-time applications. Overall, Figure 5 shows that the suggested model strikes a better compromise between efficiency and accuracy. Yield prediction performance analysis of error rate values is given in Table 5.

Table 5. Yield prediction performance analysis of error rate values

Methods	R^2 (Mean, Standard Deviation)	RMSE (Mean, Standard Deviation)	MAE (Mean, Standard Deviation)	MSE (Mean, Standard Deviation)
Graph Neural Network-Recurrent Neural Network (GNN-RNN)	93.32, 0.85	0.8872, 0.041	0.7683, 0.038	0.7871, 0.052
Multi-Convolutional Neural Network	95.01, 0.72	0.6598, 0.033	0.5714, 0.029	0.4353, 0.041
Support Vector Machine (SVM)	96.12, 0.65	0.5116, 0.028	0.443, 0.025	0.2617, 0.030
Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)	97.69, 0.54	0.3033, 0.021	0.2627, 0.019	0.092, 0.015
Proposed	**99.5, 0.21**	**0.0654, 0.006**	**0.0566, 0.004**	**0.0043, 0.001**

Note: RMSE = Root Mean Squared Error; MAE = Mean Average Error; MSE = Mean Squared Error.



(a) (b)

Figure 6. Computational efficiency analysis showing (a) training time and (b) testing time of different models evaluated on the same dataset

Table 6. Training and testing time analysis of yield prediction value

Models	Training Time (Min)	Testing Time (Sec)
GNN-RNN	33	8
Multi-CNN	30	7
SVM	26	5
CNN-LSTM	22	3
Proposed	15	1

Note: GNN-RNN = Graph Neural Network-Recurrent Neural Network; CNN = Convolutional Neural Network; SVM = Support Vector Machine; CNN-LSTM = Convolutional Neural Network and Long Short-Term Memory.

Table 7. Ablation analysis of proposed model elements

Model Configuration	Processing Time (s)	MAE	RMSE	R^2
Without HFLO with Attention	1.32	8.20	13.05	0.86
Without HFLO Feature Selection	1.40	6.85	11.25	0.90
Without Attention Mechanism	1.18	7.10	11.82	0.89
ISVR with Basic Features (No Optimization)	1.10	9.15	14.60	0.82
SVR	1.05	10.40	16.85	0.78
Proposed	**1.25**	**5.82**	**9.45**	**0.94**

Note: HFLO = Hybrid Frilled Lizard Osprey; ISVR = improved support vector regression; SVR = Support Vector Regression; MAE = Mean Average Error; RMSE = Root Mean Squared Error.

Table 8. Comparing performance with baseline models

Model	R^2	RMSE	MAE	MSE	Processing Time (s)
Explainable boosting machine	0.84	12.98	9.10	168.50	1.60
LightGBM	0.89	11.21	7.20	125.60	2.40
XGBoost	0.91	10.60	6.75	112.40	2.85
Gaussian process regression	0.86	12.25	8.45	150.20	3.10
Proposed	**0.94**	**9.45**	**5.82**	**89.30**	**1.25**

Note: RMSE = Root Mean Squared Error; MAE = Mean Average Error; MSE = Mean Squared Error.

Training and testing time analysis of yield prediction is given in Figure 6.

The computational efficiency of the assessed models in terms of training and testing time is shown in Figure 6. To guarantee consistency, the experiments were carried out using the same hardware setup and dataset. Testing time is the amount of time needed to produce predictions for data that has not yet been observed, whereas training time is the total amount of time needed to learn the model from the training dataset. Compared to GNN-RNN (33 minutes) and CNN-LSTM (about 25 minutes), the Improved SVR model takes around 15 minutes to train, as seen in Figure 6(a). In a similar vein, Figure 6(b) demonstrates that, in comparison to CNN-LSTM (3 seconds) and GNN-RNN (8 seconds), the Improved SVR achieves the fastest testing time (1 second). These results demonstrate that the proposed paradigm is perfect for real-time agricultural applications where timely decision-making is essential. The shorter training period allows for faster model updates, while the short testing duration ensures rapid yield prediction in resource-constrained or field conditions. Table 6 shows the yield prediction value analysis of training and testing times.

Each component of the proposed paradigm is demonstrated by the ablation study in Table 7. The importance of the attention mechanism in finding relevant feature dependencies is highlighted by the noticeable decline in performance that occurs when it is removed. When the HFLO-based feature selection is removed, error scores increase, indicating the presence of duplicated or less informative features. When both elements are eliminated simultaneously, the performance is at its lowest, indicating their complementary responsibilities. All things considered, the full model yields the best results, demonstrating the effectiveness of integrating ISVR, HFLO, and an attention mechanism for accurate tapioca yield prediction.

Table 8's performance comparison demonstrates that the proposed ISVR model outperforms the baseline models across all evaluation criteria. It has the lowest MAE, MSE, and RMSE values and the highest R2 score, indicating better variance explanation and prediction accuracy. While XGBoost and LightGBM (Gradient Boosting Machine) models are good at capturing spatial-temporal data, their computational cost is higher, and their generalization performance is slightly inferior. Explainable boosting machine operates steadily but very poorly because of its limited ability to model complex nonlinear relationships. The Gaussian process regression paradigm increases computational complexity without producing noticeable performance gains, despite its capacity to handle structured data. All things considered, the proposed ISVR model provides the best feasible balance between accuracy and processing efficiency, making it more suitable for real-time tapioca yield prediction.

In this study, yield data are reported in kg/ha, and all assessment metrics, including MAE and RMSE, are presented in the same unit. Because of this, the model's average forecast error is less than 1 kg/ha, demonstrating excellent prediction accuracy. Prediction errors in kg/ha are indicated by RMSE (0.654) and MAE (0.0566). Pre-processing normalization and scaling are generally too low, despite the MAE value appearing incredibly low. To improve interpretability, results are displayed in both original and normalized units. In practical terms, these low prediction errors enable farmers and agronomists to make exact decisions on resource allocation, crop management, and yield forecasting. This connection between quantitative measurements and real-world agricultural applications increases the practical significance and reliability of the recommended strategy.

4.3.1 Overfitting analysis and validation strategy

To further ensure robustness, the trials were conducted over ten different runs with different random initializations. A more reliable and statistically sound evaluation is provided by the presentation of the final results' average and standard deviation. The updated results show that the proposed model consistently maintains high performance with little variance, validating its stability. To further investigate model behavior, learning curves for training and validation loss have been introduced. Because both training and validation errors progressively decrease and converge without noticeable divergence, the charts demonstrate that overfitting is well managed.

Table 9. Statistical performance above multiple runs

Metric	Mean	Standard Deviation
MAE	0.0566	0.0042
R^2	0.995	0.0021
RMSE	0.654	0.0315

Note: MAE = Mean Average Error; RMSE = Root Mean Squared Error.

Table 10. Statistical performance over multiple runs (10 Runs)

Model	RMSE (Mean, Standard Deviation)	MAE (Mean, Standard Deviation)	R^2 (Mean, Standard Deviation)
Proposed ISVR	0.654, 0.0315	0.0566, 0.0042	0.995, 0.0021
Multi-CNN	0.750, 0.0428	0.0720, 0.0058	0.978, 0.0052
CNN-LSTM	0.720, 0.0452	0.0675, 0.0061	0.981, 0.0045
GNN-RNN	0.790, 0.0483	0.0845, 0.0069	0.970, 0.0059
SVM	0.820, 0.0510	0.0910, 0.0075	0.965, 0.0068

Note: ISVR = improved support vector regression; CNN = Convolutional Neural Network; CNN-LSTM = Convolutional Neural Network and Long Short-Term Memory; GNN-RNN = Graph Neural Network-Recurrent Neural Network; SVM = Support Vector Machine; RMSE = Root Mean Squared Error; MAE = Mean Average Error.

Regularization approaches like optimum parameter tuning in ISVR and feature selection using HFLO also help to improve generalization. After these comprehensive evaluations, the performance metrics, such as R2, were revalidated, proving that the model produces excellent prediction accuracy without exhibiting overfitting or data leakage. Table 9 shows the statistical results over several runs.

The mean and standard deviation of each model's performance over multiple runs are shown in Table 10. The proposed ISVR model has the lowest error values with the least amount of volatility when compared to existing methods, demonstrating excellent accuracy and stability.

Table 11 displays the cross-validation results, which show consistent performance across all folds. The low variation in MAE, RMSE, and R2 values indicates that the proposed model is stable and successfully generalizes to new data.

Table 11. Results of cross-validation

Fold	R^2	MAE	RMSE
Fold 1	0.994	0.058	0.662
Fold 2	0.996	0.055	0.648
Fold 3	0.995	0.057	0.651
Fold 4	0.996	0.054	0.640
Fold 5	0.993	0.059	0.670
Average	**0.995**	**0.0566**	**0.654**

Note: MAE = Mean Average Error; RMSE = Root Mean Squared Error.

5. CONCLUSIONS

In order to create a multimodal enhanced SVR framework for accurate tapioca yield prediction, this study combined IoT sensor data with UAV/satellite photography. The model's capacity to take into account both temporal and spatial factors affecting crop development contributes to its higher predicted accuracy when compared to conventional ML methods. Experimental results demonstrate that the proposed approach achieves outstanding accuracy, exhibiting great predictive capability and robustness for precision agriculture applications, with an MAE of 0.0566, RMSE of 0.654, and a high R2 value of 99.5%. Despite these positive outcomes, it's critical to acknowledge certain limitations. The model may not be as applicable to other regions with different soil and environmental conditions because the dataset was collected from a specific geographic area. Additionally, the model's effectiveness may be impacted by seasonal variations and harsh weather circumstances that were not fully examined. The reliance on high-resolution UAV and satellite pictures may pose computational and scalability challenges for large-scale or real-time deployment. Subsequent studies will focus on strengthening the model's ability to generalize by validating it across a range of geographical regions and crop types. By utilizing long-term climate data and weather forecasting information, the model's resilience to shifting environmental conditions can be further improved. Decision-making in agricultural planning may also be enhanced by taking economic and market-related factors into account. Finally, developing lightweight, real-time deployable versions of the model will facilitate its actual use in precision agricultural systems. A limitation of this work is the use of simple concatenation for multimodal fusion, which is unable to accurately capture complex interactions between sensor and picture data. Future studies will look at more advanced techniques that could improve performance, such as cross-modal attention.

REFERENCES

[1] Toromade, A.S., Soyombo, D.A., Kupa, E., Ijomah, T.I. (2024). Reviewing the impact of climate change on global food security: Challenges and solutions. International Journal of Applied Research in Social Sciences, 6(7): 1403-1416. https://doi.org/10.51594/ijarss.v6i7.1300

[2] Fuentes-Peñailillo, F., Gutter, K., Vega, R., Silva, G.C. (2024). Transformative technologies in digital agriculture: Leveraging Internet of Things, remote sensing, and artificial intelligence for smart crop management. Journal of Sensor and Actuator Networks, 13(4): 39. https://doi.org/10.3390/jsan13040039

[3] Jubair, S., Domaratzki, M. (2023). Crop genomic selection with deep learning and environmental data: A survey. Frontiers in Artificial Intelligence, 5: 1040295. https://doi.org/10.3389/frai.2022.1040295

[4] Herr, A.W., Adak, A., Carroll, M.E., et al. (2023). Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. Crop Science, 63(4): 1722-1749. https://doi.org/10.1002/csc2.21028

[5] Devi, B., Devi, J., Bhattacharyya, N. (2024). Subsistence agriculture—An approach towards food security in changing climate. In Food Production, Diversity, and Safety Under Climate Change, Springer Nature Switzerland, Springer, Cham, pp. 53-62. https://doi.org/10.1007/978-3-031-51647-4_5

[6] Borku, A.W. (2025). Cassava (Manihot esculenta Crantz): Its nutritional composition insights for future research and development in Ethiopia. Discover Sustainability, 6(1): 404. https://doi.org/10.1007/s43621-025-00996-2

[7] Oniang'o, R., Maingi, Z., Jaika, S., Konyole, S. (2025). Africa's contribution to global sustainable and healthy diets: A scoping review. Frontiers in Nutrition, 12: 1519248. https://doi.org/10.3389/fnut.2025.1519248

[8] Putri, I.A., Abidin, Z., Murniati, K. (2025). Assessing superior supply chain performance in the tapioca flour agroindustry using SCOR and FSCN models in Lampung province. Buletin Penelitian Sosial Ekonomi Pertanian Fakultas Pertanian Universitas Haluoleo, 27(1): 40-52. https://doi.org/10.37149/bpsosek.v27i1.1788

[9] Ambethgar, A.S., Rameshkumar, A., Krishna, K.R., Sundaresan, S. (2025). Biological control of pests in major tropical vegetable crops: A review. Agricultural Reviews, 46(4): 515-530. https://doi.org/10.18805/ag.R-2635

[10] Shahbazi, F., Shahbazi, S., Zare, D. (2025). Losses in agricultural produce: Causes and effects on food security. Food and Energy Security, 14(3): e70086. https://doi.org/10.1002/fes3.70086

[11] Wen, G., Cao, Y., Wei, X. (2025). The data-driven analysis of soil health and crop adaptability: Technologies, impacts, and optimization strategies. Advances in Resources Research, 5(1): 350-368. https://doi.org/10.50908/arr.5.1_350

[12] Jabed, M.A., Murad, M.A.A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with

insights for future research and sustainability. Heliyon, 10(24): e40836. https://doi.org/10.1016/j.heliyon.2024.e40836
[13] Zafar, A. (2024). Integration of IoT and machine learning for predictive analytics in smart farming: Techniques, challenges, and future directions. Journal of Advanced Computing Systems, 4(7): 1-10.
[14] Sharma, S.K., Sharma, D.P., Gaur, K. (2023). Crop yield predictions and recommendations using random forest regression in 3A agroclimatic zone, Rajasthan. Journal of Data Acquisition and Processing, 38(2): 1635-1651. https://doi.org/10.5281/zenodo.776786
[15] Rajakumaran, M., Arulselvan, G., Subashree, S., Sindhuja, R. (2024). Crop yield prediction using multi-attribute weighted tree-based support vector machine. Measurement: Sensors, 31: 101002. https://doi.org/10.1016/j.measen.2023.101002
[16] Mahesh, P., Soundrapandiyan, R. (2024). Yield prediction for crops by gradient-based algorithms. PloS One, 19(8): e0291928. https://doi.org/10.1371/journal.pone.0291928
[17] Peng, M., Liu, Y., Khan, A., et al. (2024). Crop monitoring using remote sensing land use and land change data: Comparative analysis of deep learning methods using pre-trained CNN models. Big Data Research, 36: 100448. https://doi.org/10.1016/j.bdr.2024.100448
[18] Krishna, M.V., Swaroopa, K., SwarnaLatha, G., Yasaswani, V. (2024). Crop yield prediction in India based on mayfly optimization empowered attention-bi-directional long short-term memory (LSTM). Multimedia Tools and Applications, 83(10): 29841-29858. https://doi.org/10.1007/s11042-023-16807-7
[19] Shingade, S.D., Mudhalwadkar, R.P. (2023). Hybrid extreme learning machine based bidirectional long short-term memory for crop prediction. Concurrency and Computation: Practice and Experience, 35(2): e7482. https://doi.org/10.1002/cpe.7482
[20] Ali, A., Kaul, H.P. (2025). Monitoring yield and quality of forages and grassland in the view of precision agriculture applications—A review. Remote Sensing, 17(2): 279. https://doi.org/10.3390/rs17020279
[21] Ramesh, V., Kumaresan, P. (2025). Stacked ensemble model for accurate crop yield prediction using machine learning techniques. Environmental Research Communications, 7(3): 035006. https://doi.org/10.1088/2515-7620/adb9c0
[22] Nikhil, U.V., Pandiyan, A.M., Raja, S.P., Stamenkovic, Z. (2024). Machine learning-based crop yield prediction in south India: Performance analysis of various models. Computers, 13(6): 137. https://doi.org/10.3390/computers13060137
[23] Kavitha, P.S., Nageswari, R., Kalarani, M.K., Venkatachalam, S.R., Velmurugan, M., Sudha, A. (2025). Evaluation of cassava genotypes for hilly areas under rainfed conditions. Plant Science Today, 12(3): 6557. https://doi.org/10.14719/pst.6557
[24] Prasanna, K.S., Reddy, R.A., Srilekha, M., Neelima, K., Yadav, J. (2024). Automated tuber crop harvesting using IoT and intelligent yield prediction for precision farming. In 2024 2nd International Conference on Networking and Communications (ICNWC), Chennai, India, pp. 1-7. https://doi.org/10.1109/ICNWC60771.2024.10537442
[25] Yewle, A.D., Mirzayeva, L., Karakuş, O. (2025). Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction. Remote Sensing Applications: Society and Environment, 38: 101613. https://doi.org/10.1016/j.rsase.2025.101613
[26] Shawon, S.M., Ema, F.B., Mahi, A.K., Niha, F.L., Zubair, H.T. (2025). Crop yield prediction using machine learning: An extensive and systematic literature review. Smart Agricultural Technology, 10: 100718. https://doi.org/10.1016/j.atech.2024.100718
[27] Kalmani, V.H., Dharwadkar, N.V., Thapa, V. (2025). Crop yield prediction using deep learning algorithm based on CNN-LSTM with attention layer and skip connection. Indian Journal of Agricultural Research, 59(8): 1303. https://doi.org/10.18805/IJARe.A-6300
[28] Menon, A.S., Aravinth, J., Rajendran, S., Kiran, P. (2025). Deep learning-based farm-level crop yield prediction using multi-temporal satellite data for complex engineering application. Smart Agricultural Technology, 12: 101562. https://doi.org/10.1016/j.atech.2025.101562
[29] Jiang, C., Guo, X., Li, Y., et al. (2025). Multimodal deep learning models in precision agriculture: Cotton yield prediction based on unmanned aerial vehicle imagery and meteorological data. Agronomy, 15(5): 1217. https://doi.org/10.3390/agronomy15051217
[30] Hassankhani Dolatabadi, S., Budinská, I., Behmaneshpour, R., Gatial, E. (2023). Closing the data gap: A comparative study of missing value imputation algorithms in time series datasets. In Proceedings of the Computational Methods in Systems and Software, pp. 77-90. https://doi.org/10.1007/978-3-031-53552-9_7
[31] Cabello-Solorzano, K., Ortigosa de Araujo, I., Peña, M., Correia, L., Tallón-Ballesteros, A.J. (2023). The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. In International Conference on Soft Computing Models in Industrial and Environmental Applications, Springer Nature Switzerland, pp. 344-353. https://doi.org/10.1007/978-3-031-42536-3_33
[32] Mohseni, M., Mofidi, M. (2024). Fault diagnosis of ball mills using 2D convolutional neural networks with multi-channel vibration signals. http://doi.org/10.2139/ssrn.4880712
[33] Ashfaq, N., Khan, M.H., Nisar, M.A. (2024). Identification of optimal data augmentation techniques for multimodal time-series sensory data: A framework. Information, 15(6): 343. https://doi.org/10.3390/info15060343
[34] Azawi, N. (2023). Performance comparison of deep convolution neural network and median filter in terms of denoising and detail preserving. Multimedia Tools and Applications, 82(29): 45733-45745. https://doi.org/10.1007/s11042-023-16336-3
[35] Falahah, I.A., Al-Baik, O., Alomari, S., et al. (2024). Frilled lizard optimization: A novel bio-inspired optimizer for solving engineering applications. Computers, Materials & Continua, 79(3): 3631-3678. https://doi.org/10.32604/cmc.2024.053189
[36] Dehghani, M., Trojovský, P. (2023). Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems. Frontiers in Mechanical Engineering, 8: 1126450. https://doi.org/10.3389/fmech.2022.1126450

[37] Das, S., Sahoo, B.K. (2025). Brain tumor identification using DenseNet-201 deep learning model. In Design Optimization Using Artificial Intelligence, CRC Press, pp. 193-210. https://doi.org/10.1201/9781003589716-13
[38] Panda, M.K., Subudhi, B.N., Veerakumar, T., Jakhetiya, V. (2023). Modified ResNet-152 network with hybrid pyramidal pooling for local change detection. IEEE Transactions on Artificial Intelligence, 5(4): 1599-1612. https://doi.org/10.1109/TAI.2023.3299903
[39] Ashtikar, A., Shastri, S. (2025). A CNN model for skin cancer detection and classification by using image processing techniques. Journal of Scientific Research and Technology, 3(6): 251-263.
[40] Behera, M.P., Sarangi, A., Mishra, D., Sarangi, S.K. (2023). A hybrid machine learning algorithm for heart and liver disease prediction using modified particle swarm optimization with support vector machine. Procedia Computer Science, 218: 818-827. https://doi.org/10.1016/j.procs.2023.01.062
[41] Jayanthi, S., Priya, D.T., Goud, N., Rajkumar, A., Srivalli, B. (2024). Sowing intelligence: Advancements in crop yield prediction through machine learning and deep learning approaches. https://doi.org/10.21203/rs.3.rs-4919385/v1