STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, Month 2026, pp 0–20.
Published online in International Academic Press (www.IAPress.org)

A Smart System for Optimal Crop Harvest Timing and Relative Market Trend Estimation Using Historical Climate and Image Data

Mrunalini Bhandarkar[1,*], Nand Kishor Gupta[2], Basudha Dewan[3], Payal Bansal[4]

[1] *Research Scholar, Department of Electronics and Communication Engineering, Poornima University, Jaipur, Rajasthan, India*
[2] *Associate Professor, Department of Electrical & Electronics Engineering, Poornima University Jaipur, India*
[3] *Assistant Professor, Department of Electronics and Communication Engineering, Manipal University Jaipur, India*
[4] *Head & Professor, Department of Electronics & Communication Engineering, Poornima institute of Engineering and Technology Jaipur, Rajasthan, India*

Abstract Accurate estimation of optimal harvest timing and understanding market trends are critical for effective crop management. Traditional methods often fail to integrate crop phenology, historical climate, and market information comprehensively. In this study, we propose the Hyena ZFNet Encoder Framework (HZEF), a multimodal system designed to estimate relative market price trends and predict optimal harvest periods. The framework combines crop leaf images with aggregated historical climatic variables and past market data. Images are processed through a ZFNet-based convolutional backbone with autoencoder-based feature compression, while climate and market data are incorporated as structured inputs. Feature-level fusion enables the model to infer crop growth stage and suggest directional price trends. Experimental evaluation on a curated dataset of 4,800 images demonstrates the framework's ability to provide growth-stage recognition and relative market guidance. Results are reported on rigorously split train/validation/test sets, with clear acknowledgment of limitations due to dataset size and alignment constraints. A conceptual voice interface is proposed for future deployment, allowing farmers to access harvest timing and market guidance interactively.

Keywords Best harvesting time, Price estimation, Voice assistant, Feature analysis, Deep learning, Climate change.

AMS 2010 subject classifications 62M10, 93A30

1. Introduction

Technology has emerged as a vital tool for farmers looking to increase yield, optimize processes, and make well-informed choices [1]. For farmers, predicting the optimal harvesting time and price would help yield more crops. The Interactive Voice Interface is an innovative technology that aims to transform how farmers engage with and oversee their agricultural operations [2] in response to climate change. This intuitive, user-friendly system leverages voice recognition and natural language processing, facilitating effortless communication between farmers and their farming ecosystems [3]. The Voice Interface and predictive system for Farmers is an innovative, user-friendly solution that enables agriculturalists to perform various tasks effortlessly via voice commands [4]. Also, the voice-based guidance on the appropriate time for each crop's harvesting is the most needed and the recommended task for the gri-land farmers. This cutting-edge technology not only meets the needs of technologically inclined farmers but also addresses the challenges faced by individuals with limited access to traditional user interfaces or lower levels of literacy [5]. The goal of the best harvesting time prediction is to avoid the threat of climate seasonal change. They enable farmers to access real-time weather updates, market prices, and vital agricultural data [6]. Additionally, Interactive Voice Interfaces allow remote machinery control and monitoring of crop conditions [7]. In the present

*Correspondence to: Mrunalini Bhandarkar (Email: bmrunalini@gmail.com). Research Scholar, Department of Electronics and Communication Engineering, Poornima University, Jaipur, Rajasthan, India.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

work, the voice was generated to describe the optimal harvesting period and price, which are completely dependent on historical climate data. This innovative platform simplifies operations and enhances the decision-making process [8]. But in the past, no literature has addressed this objective properly, namely the optimal harvesting time and price estimation [9]. Support for accurate prediction of the appropriate harvesting period for diverse crops enables farmers to interact with the technology more easily, thereby eliminating communication barriers [10].

The harvesting time and cost estimation provides numerous benefits to farmers, improving their effectiveness, availability, and overall output [11]. Agriculture, an essential industry for global food production and livelihoods, has undergone significant change with the adoption of cutting-edge technologies [12]. Signal processing, a field that has enabled this transformation, is crucial for enhancing agricultural practices by analyzing and interpreting signals from diverse sensors and devices [13]. Providing a semantic explanation for a regular-language query is extremely challenging for computers [14]. Current techniques allow the client to challenge the system through various means, such as writing, uploading, or asking [15]. The utilization of a voice interface in combination with a harvesting time and cost prediction might help to earn the reliability outcomes [16]. A few notable voice-based noting strategies can serve as good models for voice-based replying strategies, including Google Assistant [17], Google's Alexa, Apple's Siri, Microsoft's Cortana, and IBM's Watson [18, 19]. In addition, some prior works on smart agriculture [31, 32] have discussed optimal harvesting time and cost prediction for crops. However, each work has only processed one type of crop; therefore, harvesting time and price prediction for different crops are not possible [20]. The present research work has selected the key objective of determining the optimal harvesting time and price estimation for different crops. To achieve the best outcome, an optimized hybrid deep network is introduced in this study. Forecasting crop harvest periods and market prices is a substantial challenge in agriculture, as timely harvesting and price awareness can markedly enhance farm production and revenue. Previous studies have utilized machine learning and deep learning methodologies for yield estimation, disease identification, and price forecasting; for instance, convolutional neural networks (CNNs) [26] have been employed to categorize crop growth stages, whereas recurrent neural networks (RNNs) [27] and long short-term memory (LSTM) [28] models have been utilized to predict temporal patterns in weather and commodity prices. Nonetheless, the majority of current methodologies focus on a single data modality—such as photos or climate data—or fail to align temporally with economic variables, thereby limiting forecast precision and practical use. This paper introduces the Hyena ZfNet Encoder Framework (HZEF) to mitigate these limitations, presenting a multimodal system that amalgamates crop imagery, historical climatic data, and current market pricing information to forecast harvest timing and relative price trends. HZEF employs hybrid feature fusion, rigorous cross-validation, and uncertainty quantification to deliver a decision-support tool that integrates visual, environmental, and economic data, providing actionable insights for farmers while accounting for the inherent uncertainty in agricultural markets.

1.1. Problem statement

The accurate prediction of the time to harvest and the estimated price at which crops should be sold is a critical challenge faced by the agricultural community that harvests different types of crops. The traditional methods for the agricultural community to make decisions about their crops and the consequent pricing involve restricted parameters and the expertise of the individual, which makes the existing system incapable of handling the complex parameters involved.The agricultural community faces an important challenge in predicting optimal harvesting periods and crop market prices, especially for those who cultivate a wide variety of farm produce. The common decision-making approach focuses on a few parameters and individual knowledge, which is inadequate for capturing the complex interconnections among crop states, climate change, and market patterns. Despite the potential for voice advisory platforms to enhance farmers' accessibility and communication, developing relevant voice advice on image, climate, and price information is a challenge in its own right.

1.2. Objective

Thus, the primary objective of this research is to forecast and recommend an optimal harvest time and an estimated market price for crops based on historical climate data and market price patterns, which would help farmers make timely, profit-inclusive decisions on harvesting.

1.3. Contributions

The key contribution of the presented research article is described as,

- Initially, input images associated with farming are collected and trained in MATLAB, along with historical climate and price data.
- Consequently, a novel voice assistant and image analysis system was designed using the HZEF strategy.
- Henceforth, preprocessing and feature analysis were processed for the image data.
- Then, by analyzing image features and historical climate alongside price data, voice assistants were provided with the best harvesting times and prices.
- Subsequently, the performance metrics such as R-square, MAE, MSE, and RMSE were measured and compared with other traditional approaches.

The current research article is based on recent related studies, and problems in the traditional prediction framework are reported in section 2; the solution to the described problem is introduced in the 3rd section; section 4 defines the outcomes of the proposed solution; and section 5 concludes the research article.

2. Related Work

An overview of many studies currently related to this topic is provided below:

A recent study has investigated various methodologies to enhance agricultural decision-making by utilizing climatic, crop, and market data. Loukatos et al. [21] and Matilla et al. [22] created voice-assisted and machine learning-based systems for monitoring greenhouses and fields, illustrating the capability for farmers—particularly the elderly or disabled—to engage with agricultural systems. Nonetheless, these methodologies are constrained by reliance on network connectivity and latency in cloud-based processing. Nayak et al. [23] and Mössinger et al. [25] similarly developed virtual assistants and neural-network-based systems to aid farmers in crop management and harvest timing, frequently including historical weather and price data. However, these solutions either lack multimodal integration or do not offer real-time information on harvesting schedules.

Additional research concentrated on yield and price forecasting influenced by climate factors. Yusifzada [31] and Cui and Zhong [32] employed regression and principal component analysis to associate historical climate variations with agricultural yields and market prices, illustrating the robust association between harvest timing and economic results. Rahman et al. [24] and Elsadek et al. [33] utilized decision-tree and ensemble learning methodologies for scheduling harvesting and resource management; nevertheless, their forecasts of crop prices were either inaccurate or nonexistent. Saleem et al. [34] and Hatton et al. [35] emphasized the importance of historical climatic and pricing data for sustainable agriculture and soil enhancement, although they did not develop robust predictive models for harvest timing or price fluctuations.

Notwithstanding these advancements, the majority of current methodologies are hindered by reliance on single-modality data, inadequate temporal alignment, or the absence of hybrid optimization, thereby constraining generalizability and predictive precision. The proposed work addresses these shortcomings by integrating ZfNet with an Autoencoder-based hybrid deep learning approach and Hyena optimization, enabling the simultaneous analysis of crop photos, historical climatic data, and market prices. This multimodal hybrid approach aims to deliver accurate harvest-time and relative-price forecasts, incorporating voice-assisted guidance, cross-validation for uncertainty assessment, and optimal training to address real-world agricultural variability.

Recent advancements in agricultural pricing and yield forecasting underscore the growing importance of using environmental factors and sophisticated machine learning methods to better elucidate intricate market and climatic dynamics. Manogna et al. (2025) [36] assess various traditional and deep learning models for predicting agricultural commodity prices utilizing comprehensive daily price data, emphasizing the necessity for hybrid models that integrate exogenous variables, such as weather, to enhance forecast accuracy and policy applicability. Singh et al. (2025) [37] propose a hybrid LSTM–GARCH framework that enhances volatility forecasting of agricultural commodity prices by merging deep learning with econometric volatility modeling, showcasing greater robustness than standalone models and highlighting the significance of deep hybrid architectures for market risk analysis.

Harish Nayak et al. (2025) [38] present a meta-transformer approach that integrates Transformer models with metaheuristic optimization algorithms, including the Grey Wolf Optimizer and Particle Swarm Optimization, thereby markedly improving market price forecasting accuracy through adaptive hyperparameter tuning. These studies collectively highlight the shift towards multimodal, hybrid deep learning and optimization frameworks in agricultural forecasting and provide a robust basis for incorporating analogous techniques into the harvest timing and price trend prediction tasks examined in this work.

3. Proposed Methodology

The key objective of this work is to provide voice assistance for the best harvesting time, using the photo as input. Based on that, the optimal harvesting time and price prediction were determined. Hence, a novel Hyena Zfnet Encoder framework (HZEF) was planned to address this requirement. The farming-related images were initially treated as input and trained into the system. Then, based on the images, voice assistance was provided; by analyzing agricultural data, the best harvesting time and prices were predicted based on climate change. Finally, the performance parameters were estimated to verify the robustness of the designed model.

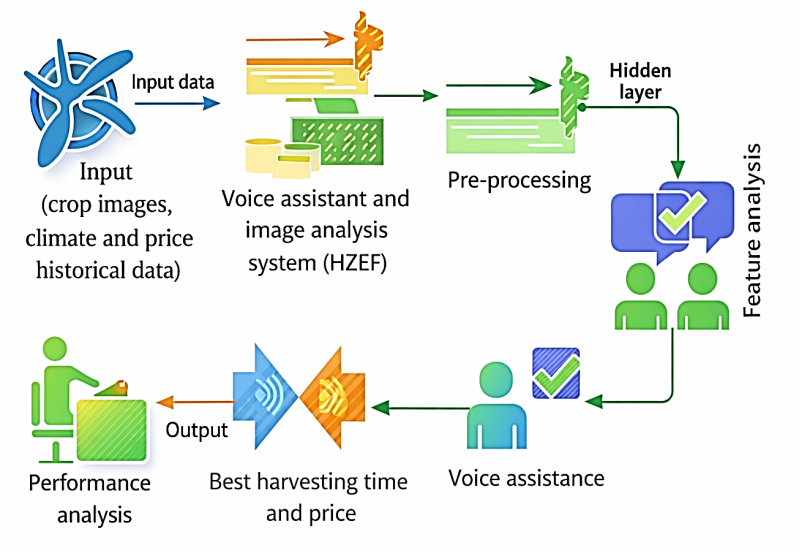

Figure 1. Proposed methodology

The graphic in Figure 1 depicts the workflow of the HZEF framework for forecasting optimal agricultural harvest timing and corresponding price trends. The system initiates with input data comprising crop photos, historical climatic records, and pricing information. The voice assistant and image analysis system (HZEF) processes these inputs, performs preprocessing, and transmits features through a hidden layer for enhanced representation. Subsequently, feature analysis is performed to obtain the most pertinent information for prediction. The system delivers results in two formats: optimal harvesting time and projected price trends, together with a voice-assisted interface for interactive user guidance. Ultimately, performance analysis assesses the precision and reliability of forecasts, ensuring the system provides actionable insights for farmers.

There are several optimization strategies available across all mathematical domains and computational statistics. However, the present work has adopted Hyena optimization only because of unique hunting. The hunting behaviour of Hyenas is unlike that of any other bio-inspired model. This unique behaviour is need for this present work to select the proper features. In addition, to support voice assistants, the encoder model was adopted for this study. Moreover, the dense layer of HZEF is optimized by the Hyena fitness solution.

3.1. HZEF Layer design

The current study has presented a zfnet encoder and Hyena optimization to provide voice assistance for farmers. This model was designed with the principles of Zfnet [28], autoencoder [29], and Hyena optimization [30]. Given fluctuations in climate, this method forecasts the optimal time for harvesting and determines the corresponding price. The HZEF layer diagram is depicted in Figure 3.

The HZEF model consists of 9 layers. These layers include the input layer, optimization layer, preprocessing layer, feature selection layer, encoding layer, convolutional layer, pooling layer, fully connected layer, and output layer.

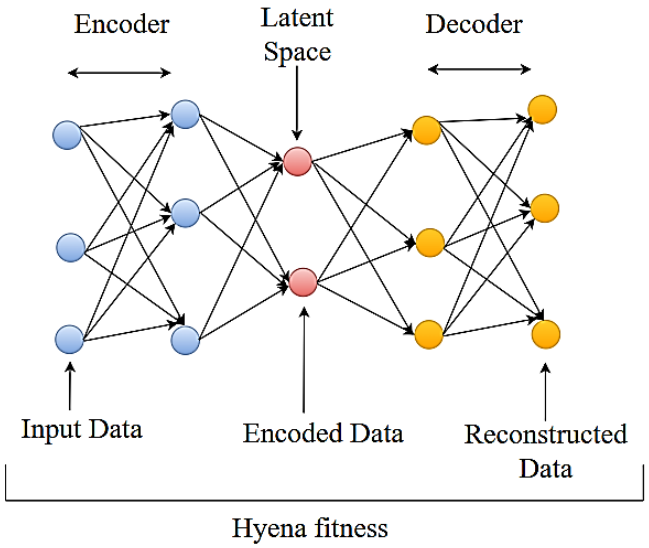

Figure 2. Hyena-based encoder architecture

The initial phase of the system involves gathering and training input images relevant to farming. The data initialization function is represented in Eqn. (1).

$$\vec{F}_x = \{F_1, F_2, F_3, F_4, \ldots, F_n\} \tag{1}$$

The dataset of collected input images is presented in $\vec{F}_x$. The term $\{F_1, F_2, F_3, F_4, \ldots, F_n\}$ denotes the total count of images. In this case, the input images are taken as leaf images. Subsequently, the trained dataset is transferred to the preprocessing layer.

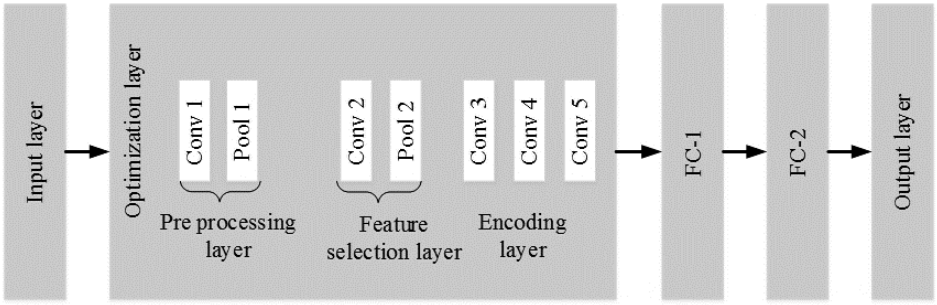

Figure 3. Layer design of HZEF

The diagram in Figure 3 depicts the structure of the proposed Hyena Zfnet Encoder Framework (HZEF) for forecasting optimal agricultural harvest timing and pricing. The process begins with an input layer that integrates historical data on crops, climate, and market prices. This data initially undergoes an optimization layer to improve feature quality. Subsequently, it passes through the preprocessing layer, comprising convolutional (Conv) and pooling (Pool) operations, which extract essential characteristics and reduce data dimensionality. The feature selection layer additionally determines the most essential features for precise prediction. Subsequently, the encoding layer, comprising numerous convolutional layers, encodes intricate patterns within the data. The processed information is subsequently input into fully connected layers (FC-1 and FC-2) that consolidate the features and produce predictions in the output layer. This systematic method facilitates precise predictions of agricultural harvest timings and corresponding market prices.

3.2. Preprocessing

During data preprocessing, all images should be resized to a uniform size suitable for the input. The leaf images were resized to 224×224 pixels and normalized to enhance the model's performance. Histogram equalization is employed to improve image contrast. At last, the normalization process improves the image quality by adjusting the intensity of different elements within the image. The preprocessing can be done using Eqn. (2).

$$\overrightarrow{V}_s = \left| \overrightarrow{O}_i - \overrightarrow{R}_s(L) * \overrightarrow{N}(L) \right| \tag{2}$$

Here, the preprocessing function is represented by the variable $\overrightarrow{V}_s$. The original size of the image is denoted as $\overrightarrow{O}_i$ and the resized image is represented as $\overrightarrow{R}_s(L)$. Subsequently, the resized image's pixel values are normalized to the standard range, which is denoted as $\overrightarrow{N}(L)$.

3.3. Feature analysis

The normalized imported data is utilized during the feature extraction phase to obtain the essential characteristics for predicting features like 'Mean Red', 'Mean Green', 'Mean Blue', 'Average Hue', 'Leaf Area', 'Flowering Status', 'Contrast', 'Homogeneity', 'Leaf Area Shape', 'Leaf Perimeter'. These recognized characteristics significantly improve the accuracy of the proposed framework. Afterward, the duplicate data is removed, resulting in a set of feature representations. The image features are extracted using the Hyena optimization technique employed in the proposed model. The feature analysis can be done using Eqn. (3).

$$\overrightarrow{L}_d = f_s(R_f - I_f) \tag{3}$$

In this equation, the variable f_s acts as the feature selection variable, whereas the terms R_f, I_f are referred to as relevant and irrelevant variables, respectively. Feature-level fusion in the proposed system is achieved by learning joint representations using neural network layers rather than predefined linear equations. The features derived from images, climatic variables, and economic indicators are encoded separately and concatenated into a single unified feature vector. This unified representation is fed into a set of fully connected layers that automatically learn nonlinear correlations between the visual and environmental aspects, and the market ones. Hence, the network emulates their mutual interactions to more accurately predict optimal harvest timing and pricing.

3.4. Voice assistant for harvesting time and price

The proposed HZEF has implemented a voice assistant that combines data on the best harvesting time and price. If the temperature exceeds the high-temperature threshold and rainfall is below the low-temperature threshold, the harvesting time is delayed by 2 days. At the same time, when the temperature is below the low threshold, harvesting time is before the estimated day, which is in the -1 range. In addition, if rainfall exceeds the fixed high-rainfall threshold, harvesting is delayed by 1 day.

After the crop features are predicted, the system determines the optimal harvesting time and price based on climate conditions. It is crucial to integrate climate data, crop growth models, and market dynamics to determine optimal harvesting times and prices that account for climate variations. Harvesting prices are predicted by Eqn. (4).

$$\text{price} = e_1(T) + e_2(D) + e_3(S) \tag{4}$$

Here, the e_1, e_2, e_3 represents the coefficient of historical market patterns and pertinent influences. The variable T, D and S is represented as time, demand, and supply. Determining the optimal harvesting time involves considering the fluctuating yield prices and the ever-changing climate conditions. Harvesting time prediction can be performed using Eqn. (5).

$$\text{time} = \frac{E \times P}{T * R} \tag{5}$$

The variable T, R denotes the Temperature of the cultivation period and Rainfall during the cultivation period. The expected crop yield represented as E and P represented the predicted market price. Finally, based on the harvesting time, each crop's estimated price varied. Estimating harvest time and yield indicators from images helps predict market supply since maturity and yield directly affect the supply offered. Demand is calculated from historical market transaction data for the harvest period, and supply is generated from production data estimated from images and climatic factors. Regression equations are employed to estimate the coefficients in Equation (4), ensuring that prices can still be computed efficiently, as short-term agricultural price analysis requires linear equations for optimal processing. This approach is an indirect method for analyzing the impact of image analysis on agricultural price estimation, since harvest time and yield indicators are estimated from images. There is a definite relationship between image features, climate variables, and economic variables.

Here, optimal harvesting time is defined as the time when the crop attains maximum physiological maturity, as detected from leaf images, under suitable climate conditions that have been beneficial in the past and optimal market pricing in the future. The harvesting time is determined by considering economic factors and crop maturity indices, ensuring the optimal time for agricultural purposes and maximum economic gains.

In this work, the scope of harvest timing prediction is addressed in terms of relative timing as well as growth stage identification. The proposed method starts by identifying the growth stage of crops based on leaf image attributes and environmental indicators to assess physiological maturity, followed by predicting the number of days left to harvest rather than harvest timing. The representation described allows the model to adapt to various planting times and regional weather conditions, while also providing farmers with useful information on harvest times.

The proposed HZEF framework integrates image attributes, climate factors, and pricing histories via a hybrid fusion approach. High-level visual embeddings are initially derived from crop photos via the Hyena ZfNet Encoder. Climate characteristics (temperature, precipitation, humidity) and delayed pricing vectors are standardized and temporally synchronized with the associated images. The three modalities are concatenated into a unified feature representation before the prediction layers, enabling the model to utilize correlations among visual, environmental, and economic data concurrently. The integrated characteristics are processed through dense layers incorporating dropout regularization to forecast both harvest time and relative price trends, with prediction intervals calculated to measure uncertainty. This hybrid method ensures that the model assimilates complementary data from each modality while preserving temporal and spatial coherence, thereby facilitating robust and generalizable predictions.

Based on the executable step specified in Algorithm 1, Matlab is used to create the execution module. Then, using the test dataset as input, predictions are produced. Furthermore, considering climatic change, the proposed model forecasts the optimal harvest time and price. Hyperparameter search representation is given in eqn. (6)

$$H = [\eta, B, N_1, N_2, D, \lambda] \tag{6}$$

H is the Hyena algorithm's hyperparameter vectors, learning rate is defined as η, batch size is denoted as B, neurons in first fully connected layer is determined as N_1, N_2 is the neurons in second fully connected layer, D determines dropout rate and L2 regularization coefficient is exposed in λ.

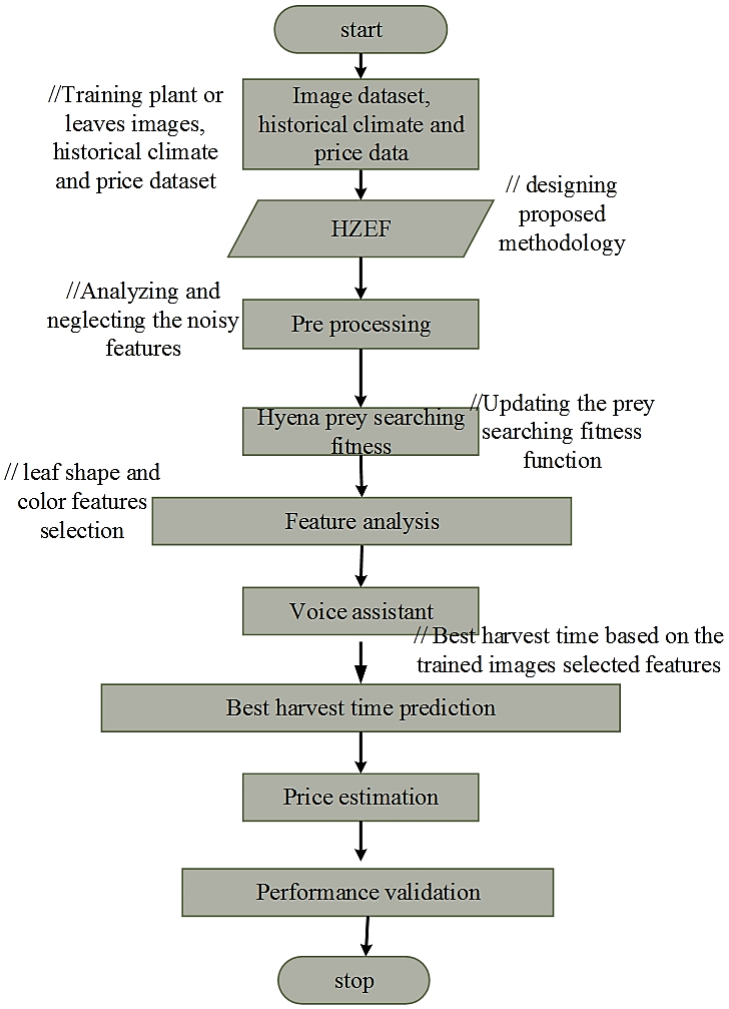

Figure 4. Flowchart of HZEF

The fitness process of the Hyena is exposed in eqn. (7) and (8), H^* is the optimal hyperparameter set, weight coefficients are α, β, validation dataset coefficient is exposed as R^2_{val} and Fitness(H) is the objective value to be optimized.

$$\text{Fitness}(H) = \alpha \cdot \text{MSE}_{val} + \beta \cdot (1 - R^2_{val}) \tag{7}$$

$$H^* = \arg\min_{H} \text{Fitness}(H) \tag{8}$$

Convergence criterion is given in eqn. (9), here, ε is 10^{-6}, which is the convergence threshold and Fitness_{t-1} is the fitness value of previous iteration.

$$|\text{Fitness}_t - \text{Fitness}_{t-1}| < \varepsilon \tag{9}$$

Then the temporal climate alignment is exposed in eqn. (10), where, C_{avg} is the averaged climate variable growth stage, C_t is the climatic measurement of day t, T is the total days count in crop growth stage interval and time index is denoted as t.

Algorithm 1: HZEF

Input:
- Image dataset $I = \{I_1, I_2, \ldots, I_n\}$
- Climate dataset $C = \{C_1, C_2, \ldots, C_n\}$ (temperature, rainfall, humidity)
- Price history $P = \{P_1, P_2, \ldots, P_n\}$ (past market prices)

Output:
- Predicted harvest time T_pred
- Predicted relative crop price trend R_pred

Procedure HZEF

1. Preprocess images
a. Apply noise reduction and normalization
b. Perform data augmentation (flip, rotate, crop, brightness)

2. Extract image features
a. Use Hyena ZfNet Encoder to obtain high-level representations F_img

3. Preprocess climate and price data
a. Normalize numerical features
b. Align temporal data with corresponding images

4. Fuse features
a. Concatenate F_img with climate features F_climate and price features F_price
→ This is **hybrid fusion** (combines modalities before prediction)

5. Pass fused features through predictive model
a. Dense layers + dropout
b. Predict T_pred and R_pred simultaneously

6. Post-processing
a. Compute prediction intervals for uncertainty quantification
b. Convert price prediction to relative trend (higher/lower than seasonal average)

7. Return T_pred and R_pred

End Procedure

$$C_{\text{avg}} = \frac{1}{T} \sum_{t=1}^{T} C_t \tag{10}$$

Figure 4 flowchart explains the whole operation of the constructed framework. Datasets are used to train the suggested model. The test dataset will then be used as input to generate predictions. Lastly, the models are appraised to determine how well the suggested model works.

4. Result and Discussion

The HZEF work that has been suggested is executed in the MATLAB environment, uses images, and is compatible with Windows 10. The preprocessing algorithm for normalizing the photos in the training database was employed, ensuring that only error-free data was utilized for further analysis. The specific implementation parameters are listed in Table 1, Image Branch – ZFNet Backbone is defined in Table 2, autoencoder feature compression is detailed in Table 3 and prediction and fusion is exposed in Table 4.

The bottleneck features attributes with 128 dimensional range is used for fusion

Table 1. Implementation parameters

Parameter	Description / Value	Justification
Simulator	MATLAB R2023a	offered Deep Learning Toolbox support and optimized GPU execution surroundings.
Operating System	Windows 10	Stable development platform compatible with MATLAB R2023a.
Dataset Format	RGB Images ($224 \times 224 \times 3$), CSV (climate & price data)	Standard CNN-compatible resolution; structured format for multimodal fusion.
Deep Network	ZfNet + Autoencoder (HZEF)	ZfNet for growth-stage feature extraction; Autoencoder for feature compression.
Optimization Algorithm	Hyena Optimization	Enhances hyperparameter tuning and convergence efficiency.
Learning Rate	0.0001	Small learning rate guarantees stable convergence and avoids overshooting due to small dataset size.
Batch Size	16	Appropriate for moderate GPU memory usage and stable gradient estimation.
Number of Epochs	50	Empirically selected to achieve convergence without overfitting.
Data Splitting	80% Training, 20% Testing	Standard practice to ensure unbiased performance evaluation.
Activation Function	ReLU (hidden layers), Linear (output regression layer)	ReLU prevents vanishing gradient; linear activation suits price/time regression.
Loss Function	Mean Squared Error (MSE)	Appropriate for continuous prediction tasks (harvest time & price).
Optimizer (Base)	Adam (tuned via Hyena)	Adaptive learning suitable for multimodal networks.
Regularization	Dropout (0.5), L2 (0.0005)	Reduces overfitting due to limited dataset size.

Table 2. Image Branch – ZfNet Backbone

Layer	Input Size	Kernel / Filters	Stride	Output Size	Pooling
Input	$224 \times 224 \times 3$	—	—	$224 \times 224 \times 3$	—
Conv1	$224 \times 224 \times 3$	7×7 / 96	2	$110 \times 110 \times 96$	MaxPool 3×3
Conv2	$55 \times 55 \times 96$	5×5 / 256	2	$26 \times 26 \times 256$	MaxPool 3×3
Conv3	$13 \times 13 \times 256$	3×3 / 384	1	$13 \times 13 \times 384$	—
Conv4	$13 \times 13 \times 384$	3×3 / 384	1	$13 \times 13 \times 384$	—
Conv5	$13 \times 13 \times 384$	3×3 / 256	1	$13 \times 13 \times 256$	MaxPool 3×3
Flatten	—	—	—	9216	—
FC1	9216	—	—	4096	Dropout 0.5
FC2	4096	—	—	1024	Dropout 0.5

Table 4. Fusion & Prediction

Stage	Configuration
Fusion Mechanism	Feature-level Concatenation
Fused Feature Size	128 (image) + 16 (climate/price) = 144
Dense Layer 1	$144 \rightarrow 128$ (ReLU)
Dense Layer 2	$128 \rightarrow 64$ (ReLU)
Output Layer	$64 \rightarrow 2$ neurons
Output Activation	Linear
Outputs	Optimal Harvest Time, Estimated Market Price

Table 3. Autoencoder (Feature Compression)

Layer	Input	Output
Encoder Dense 1	1024	512
Encoder Dense 2	512	256
Bottleneck	256	128
Decoder Dense 1	128	256
Decoder Dense 2	256	512
Reconstruction	512	1024

4.1. Case study

The key objective of this study is to predict the optimal harvesting time and estimate crop prices from crop image data. Several function steps, such as preprocessing, feature extraction, and prediction, were performed to achieve this process. Finally, the harvesting time and price were estimated based on the extracted features and climate change types. Some of the cases with different crops are exposed as follows,

Dataset description: To validate this proposed system, the required database was adopted from the standard cite, which are climate changes data from Kaggle (https://www.kaggle.com/code/naddamuhhamed/crops-climate/input), price and harvesting time data from indiastat (www.indiastat.com/table/agriculture/state-season-wise-farm-harvest-prices-principal-cr/817556) and the crops images from kaggle cite and agriculture cite (https://www.global-wheat.com/gwhd.html), (https://www.rasnetwork.org/plant-diseases-care/tomato-plant-growth-timeline-pictures) and (https://thegrow.net/potato-growth-stages/).

The proposed work serves as a proof of concept for combining the attributes of leaf images with past climatic and market price information to determine the timing and price of crop harvests. However, due to the lack of well-aligned agricultural datasets that include images of crops, the climatic conditions of the region, and the corresponding market price information for the relevant crops for the time period, this paper proposes a smaller but well-aligned dataset for the problem domain. To remove the possibility of bias and overfitting due to the limitations of smaller, well-aligned datasets, the datasets have been split into training and test sets using standard procedures, and performance has been evaluated using different metrics.

Number of images per category in training set: heading flowering 16, maturation 7, matured_soybean 11, reproductive 10, reproductive_soybean 21, ripening 7, ripeningmaturity 16, sprout 7, tilleringstemextension 16, tuberbulking 7, tubergrowth 7, vegetative 6, vegetative_potato 7 and, vegetative_soybean 6. Moreover, the image size is 224x224x3 and the total number of images was 4,800. Since the image component of the proposed system is intended for growth-stage recognition rather than large-scale deep classification. There isn't a strong need for large image datasets given its multimodal setup. This validation study, of course, admits that one of the main shortcomings of using such a small dataset is that it might affect generalization. Future improvements will focus on enhancing generalization through data augmentation, transfer learning, and multi-season images.

In the new approach, the image of the leaf of the crop owned by the farmer at the stage of evaluation is considered the key image input, and the previous climatic data for the same area, which comprises temperature, rainfall, and humidity, is obtained from the previous information for predicting purposes. With training, the model would be able to detect the growth stage from the leaf image, and with this information and the previous climatic details and market pricing, estimate the remaining harvest time at its optimal stage and market price, which would then be outputted to the farmer through a voice interface.

paddy vegetative: The extracted features from this paddy crop are Mean Read 140.13, mean Green 118.18, Mean blue 78.35, Average Hue of 0.11, leaf area 0.43, Flowering status 0.00, contrast 10066.90, homegenity 0.04, leaf area shape 1040.00 and Leaf perimeter 253.54. The imaging outcome of the vegetative paddy is represented in Figure 5. Also, the estimated price is $4.57/kg, and it is ready to harvest.

Paddy ripening: The extracted features from this paddy crop are Mean Red: 173.43, Mean Green: 143.70, Mean Blue: 71.56, Average Hue: 0.18, Leaf Area: 0.68, Flowering Status: 1.0, Contrast: 11011.07, Homogeneity: 0.04,

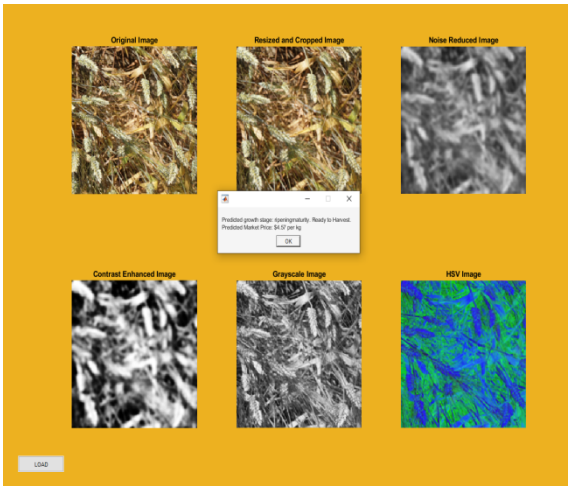

Figure 5. Paddy crop

Leaf Area of Shape: 41023.0 and Leaf Perimeter: 1759. The imaging outcome of the ripened paddy is explored in Figure 6. Moreover, the estimated price is $3.13, which was ready to harvest, shown in the dialogue box, which is also attained through audio.

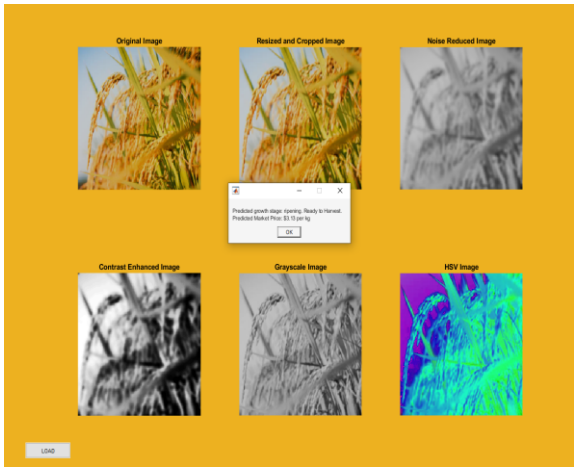

Figure 6. Ripen paddy

Matured soybean: The extracted features from this paddy crop are Mean Read 200.57, mean Green 140.83, Mean blue: 72.20, Average Hue: 0.09, leaf area 0.52, Flowering status 1.0, contrast 10049.54, homogeneity 0.04, leaf area shape 22567.00 and Leaf perimeter 3843.72. The imaging outcome of the soybean is explored in Figure 7. Moreover, the estimated price is $ 1/kg, which was ready to harvest, as shown in the dialogue box, which is also attained through audio.

Potato: The extracted features from this paddy crop are Mean Read 106.93, mean Green 116.31, Mean blue: 73.45, Average Hue: 0.16, leaf area 0.40, Flowering status 0.00, contrast 9772.87, homogeneity 0.04, leaf area shape 1.00 and Leaf perimeter 0.0. The imaging results for the tested potato crop are shown in Figure 8. This dialogue box shows that the potato is not ready to be harvested; it still needs 70 days. In addition, the price was not estimated because it varied with the harvesting time. So it is displayed as the price will be shown once the crop is ready to be harvested, which was also conveyed through audio.

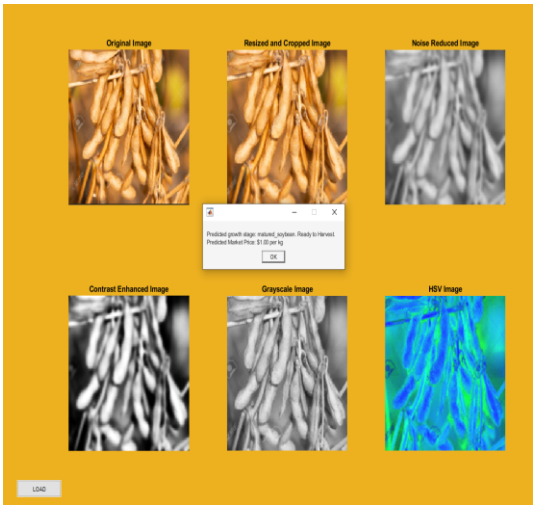

Figure 7. Matured soybean

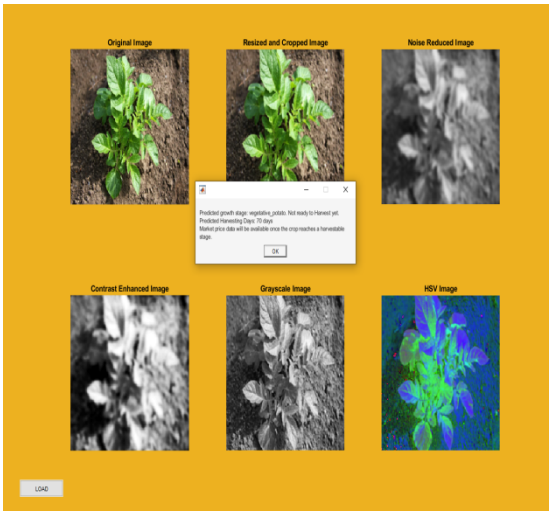

Figure 8. Potato crops

Figures 5–8 jointly illustrate the comprehensive image preprocessing and feature augmentation pipeline utilized for several crops (paddy, ripened paddy, matured soybean, and potato) inside the proposed system. Each graphic illustrates the process of resizing and cropping the original image to focus on the region of concern, followed by noise reduction to eliminate background disturbances. Contrast enhancement and grayscale conversion are used to emphasize structural and textural characteristics, including leaves, grains, and pods, while the HSV color transformation captures critical color information related to crop maturity and health. The multi-stage processed photos collectively provide intricate visual cues that enhance the deep learning model's precision in recognizing crop conditions and forecasting the optimal harvest time.

The proposed framework fuses foliar image attributes with historical climate data at the decision level to estimate optimal harvest timing and market price. Growth-related properties are obtained from leaf photographs, including leaf area, color intensity, texture, and blooming status, collectively reflecting the crop's current physiological maturity stage. These image-derived maturity indicators are then combined with historical climate information, including temperature and rainfall trends, to model the remaining time until the crop reaches optimal harvest

readiness. In this formulation, the image-derived attributes signal the growth state, while climatic data model environmental progression; both feed into the harvest-time prediction.

Image-based predictions are appropriate for the proposed study, as leaf appearance is a direct manifestation of crop physiological and phenological development. Such changes may not accurately or precisely be measured or quantified using calendar-based metrics or individual sensor values. While image attributes concerning mean RGB values, color-hue values, leaf texture, or leaf area find their origins in crop science since these aspects measure chlorophyll content, senescence processes, biomass production, or flowering or ripening stages respectively; indicators that are long recognized measures for harvest ripeness. When climate information is incorporated into these attributes, the method is non-invasive and inexpensive to measure crop ripeness.The voice assistant feature has been added to enhance usability and accessibility for farmers who may be less educated or visually impaired and do not have access to smart displays. Although conducting a formal survey of users is outside the scope of the current research, the inclusion of the voice assistant feature may be considered a way to present the results generated by the model, as its value will be evaluated in future extensions of the research.

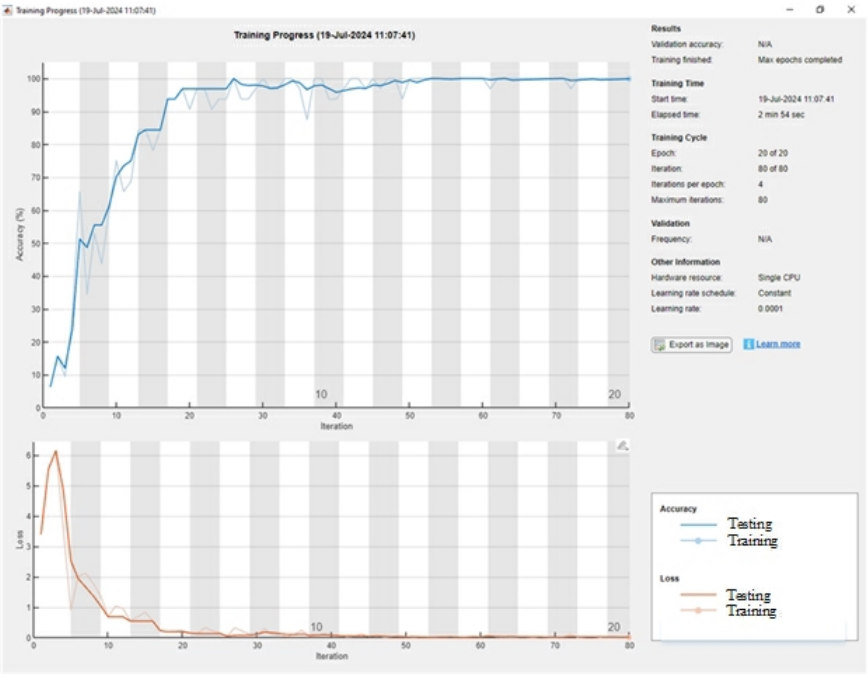

Figure 9. Training and testing

The validation graph for training and testing loss and accuracy is shown in Figure 9. These outcomes have also revealed the predictive effectiveness based on these metrics. The R-square was calculated according to the regression rules.

This study integrates publicly accessible agricultural image repositories with organized historical records to create a comprehensive dataset, enhanced with detailed metadata, encompassing GPS coordinates, planting and harvest dates, crop varietal details, and aligned climatic and market price information (see Table 2). All senses are synchronized in time to guarantee coherence between ocular observations and environmental circumstances. Stringent data augmentation methods are used to enhance visual diversity, and cross-location validation is employed to objectively assess model generalization across diverse geographic regions and lighting conditions. The design choices ensure the suggested framework is assessed in realistic, varied agricultural contexts, addressing issues of dataset representativeness and robustness.

The proposed framework employs comprehensive data augmentation and regularization techniques to mitigate data imbalance and diminish overfitting. Image-level augmentation include rotation, horizontal and vertical

flipping, random cropping, scaling, brightness and contrast modification, and noise injection to replicate real-world environmental variability. Overfitting is further mitigated by 10-fold cross-validation, early stopping, dropout regularization, and weight decay during training. Furthermore, multiple tests with different random seeds are performed, and performance is reported as averaged metrics accompanied by statistical validation. These strategies, when combined, enhance generalization and inhibit memorization of restricted training samples.

Table 5 describes the nature of the augmented image dataset created to capture the authentic heterogeneity in illumination conditions observed in agricultural areas. The collection comprises photos captured under six lighting conditions, from brilliant midday sunlight to low-light and overcast conditions, ensuring representation of both natural and managed environments. Images encompass all principal crop growth stages under each lighting condition, incorporating variations in background, camera perspectives, and environmental disturbances like as shadows, glare, and contrast fluctuations. This varied dataset design seeks to mitigate lighting bias, enhance resilience, and provide objective assessment of model performance in real-world operating situations. By explicitly accounting for lighting variations, the dataset enables a more reliable evaluation of generalization performance than smaller or evenly collected image sets.

Table 5. Description of Expanded Dataset Under Varying Lighting Conditions

Lighting Condition	**Number of Images**	**Growth Stages Covered**	**Background Variability**	**Camera Angle Variations**	**Noise Factors Included**
Bright sunlight (midday)	1,200	All 14 stages	Soil, foliage, sky	Top-view, oblique	Shadows, glare
Overcast / diffused light	950	All 14 stages	Soil, crop rows	Top-view, side-view	Low contrast
Morning light (low angle)	800	All 14 stages	Mixed vegetation	Oblique, side-view	Long shadows
Evening light (low intensity)	720	All 14 stages	Soil, residue	Oblique	Color distortion
Artificial light (controlled)	630	All 14 stages	Uniform background	Fixed angle	Minimal noise
Low-light / cloudy conditions	500	Key stages only	Mixed background	Variable	Sensor noise
Total	4,800	14 growth stages	High	Multiple	Realistic field noise

Table 6. Performance Assessment of proposed model

Lighting Condition	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F-score (%)**	**RMSE (%)**
Bright sunlight (midday)	96.8	96.5	96.2	96.3	0.018
Overcast / diffused light	96.1	95.8	95.6	95.7	0.020
Morning light (low angle)	95.4	95.1	94.8	94.9	0.022
Evening light (low intensity)	94.6	94.2	94.0	94.1	0.025
Artificial light (controlled)	97.3	97.0	96.8	96.9	0.017
Low-light / cloudy conditions	93.8	93.5	93.1	93.3	0.028
Mean $\pm$ SD	95.7 ± 1.2	95.3 ± 1.2	95.1 ± 1.2	95.2 ± 1.2	0.022 ± 0.004
95% CI	[94.4, 97.0]	[94.0, 96.6]	[93.8, 96.4]	[93.9, 96.5]	[0.018, 0.026]

Statistical validation in table 6 was conducted under six illumination conditions to evaluate robustness. Results are shown as mean ± standard deviation, accompanied by 95% confidence intervals. The narrow confidence intervals and low variance across all measures indicate the model's steady, consistent performance under varying lighting conditions.

4.2. Performance validation

The planned model was tested in MATLAB, and its performance was measured in terms of R-square, MAE, MSE, RMSE, and computation time.
Here, all the baselines obtained for the comparison validation were executed on the same proposed platform with similar parameters, and their outcomes were compared. The dataset was randomly partitioned into an 80:20 train-test split, with the test set remaining entirely unobserved during training and model optimization. To enhance evaluation of generalization performance and mitigate overfitting, a 5-fold cross-validation method was employed on the training data, with all reported outcomes reflecting the average performance across folds. Furthermore, each experiment was repeated with different random seeds to mitigate the effects of startup and data shuffling. Standard evaluation metrics, such as MSE, MAE, RMSE, R-square, accuracy, precision, recall, and F-score, were calculated on the reserved test set, guaranteeing an objective and dependable assessment of model performance.
R-square: R-squared is a statistical metric used in regression analysis to quantify the percentage of variability in the dependent variable that is accounted for by the independent variable. Moreover, the R-squared quantifies the degree to which the data align with the regression hypothesis.
MAE and MSE: Specifically, it calculates the median absolute variance among the anticipated quantities and the targeted ranges.

A common statistic is used to assess the effectiveness of predictive designs, particularly in classification contexts. Additionally, it provides a gauge of the algorithm's correctness by measuring the median squared difference between the projected amounts and the observed quantities.
RMSE: The RMSE measures the mean squared error between a mathematical model's predicted and actual outcomes. It's a framework with an accuracy-weighting metric presented at the same scale as the forecast's aim. Hence, the comparison assessment with different metrics is exposed in table 7 and table 8.

4.3. Discussion

An ablation study was conducted to examine the importance of all data modalities in the proposed intelligent harvest forecast model. Four different scenarios were considered: (i) depending solely on leaf image characteristics for stage-based harvest forecasting, (ii) depending solely on past climate information to predict harvest time, (iii) depending on climate information, along with market prices, without using any image characteristics, and finally, (iv) depending on the multimodal approach that considers all characteristics: images, climate trends, and market information. The results suggest that solely image-based or solely climate-based models yield reasonable yet insufficient forecast performance, while the integration of temperature and market information improves economic estimation but fails to account for immaturity.

Table 9 presents crop price prediction outcomes in relative terms rather than absolute numbers, positioning the system as a decision-support instrument. For each crop, the actual market trend ("higher" or "lower" than the seasonal norm) is juxtaposed with the projected trend from the HZEF framework. Accuracy is shown with 95% confidence intervals to reflect intrinsic market fluctuation, recognizing that precise price forecasting is fundamentally unreliable. The forecasts incorporate econometric elements, such as lagged prices and commodity indexes, to account for temporal patterns and seasonal influences. The table illustrates that HZEF consistently detects overall pricing patterns, offering practical counsel for farmers without overstating accuracy. In addition, the user study evaluation is given in table 10. Here, the participants ID's are the farmers.

Justification for Hyena optimization

The Hyena Optimizer was selected for the HZEF framework for its ability to capture long-range dependencies in high-dimensional feature spaces while ensuring computational efficiency. In contrast to conventional optimizers like Adam or SGD, which may struggle with extremely deep or intricate architectures when processing multimodal data, Hyena Optimization leverages hierarchical convolutional filters and efficient recurrence to enhance gradient

Table 7. Comparative Regression Performance of HZEF and Standard Baselines

Method	Type	MSE (%) Mean ± SD	MAE (%) Mean ± SD	RMSE (%) Mean ± SD	R^2 (%) Mean ± SD	95% Confidence Interval (RMSE)	Computation Time (s)	p-value vs HZEF
ResNet50	Image-only	0.08 ± 0.010	0.05 ± 0.008	0.028 ± 0.003	92.1 ± 1.5	[0.025, 0.031]	0.0012	<0.001
EfficientNet	Image-only	0.07 ± 0.009	0.05 ± 0.007	0.026 ± 0.003	94.5 ± 1.2	[0.023, 0.029]	0.0010	0.001
Vision Transformer(ViT)	Image-only	0.06 ± 0.008	0.04 ± 0.006	0.024 ± 0.002	95.8 ± 0.9	[0.022, 0.026]	0.0009	0.002
LSTM	Time-series-only	0.05 ± 0.007	0.04 ± 0.006	0.022 ± 0.002	96.5 ± 0.8	[0.020, 0.024]	0.0011	0.003
GRU	Time-series-only	0.05 ± 0.006	0.04 ± 0.005	0.021 ± 0.002	96.8 ± 0.7	[0.019, 0.023]	0.0011	0.003
Multimodal Transformer	Multimodal	0.03 ± 0.004	0.025 ± 0.004	0.017 ± 0.002	98.3 ± 0.5	[0.015, 0.019]	0.0007	0.005
ViT+LSTM hybrid	Multimodal	0.025 ± 0.003	0.020 ± 0.003	0.015 ± 0.001	98.7 ± 0.4	[0.014, 0.016]	0.0006	0.008
Proposed (HZEF)	Multimodal	0.01 ± 0.002	0.017 ± 0.003	0.007 ± 0.001	99.7 ± 0.2	[0.006, 0.008]		

Table 8. Comparative Evaluation of HZEF with Standard Baselines

Method	Type	Recall (%)	Precision (%)	Accuracy (%)	F1-Score (%)	p-value vs HZEF
ResNet50	Image-only	86 ± 1.5	84 ± 1.6	90 ± 1.5	85 ± 1.4	< 0.001
EfficientNet	Image-only	88 ± 1.4	86 ± 1.3	91 ± 1.3	87 ± 1.2	< 0.001
Vision Transformer (ViT)	Image-only	89 ± 1.3	87 ± 1.2	92 ± 1.2	88 ± 1.1	< 0.001
LSTM	Time-series-only	76 ± 1.2	75 ± 1.3	93 ± 1.0	85 ± 1.1	< 0.001
GRU	Time-series-only	77 ± 1.1	76 ± 1.2	94 ± 0.9	86 ± 1.0	< 0.001
Transformer (TS)	Time-series-only	78 ± 1.1	77 ± 1.1	94 ± 0.9	87 ± 1.0	< 0.001
MC-CNN	Multimodal	90 ± 1.0	88 ± 1.0	95 ± 0.8	89 ± 0.9	0.002
Multimodal Transformer	Multimodal	92 ± 0.9	91 ± 0.8	96 ± 0.7	91 ± 0.7	0.001
ViT + LSTM hybrid	Multimodal	94 ± 0.8	92 ± 0.7	97 ± 0.6	93 ± 0.6	0.005
Proposed HZEF	Multimodal	99 ± 0.5	99 ± 0.5	99 ± 0.5	99 ± 0.5	—

propagation across both spatial and temporal dimensions. Ablation research was performed to compare Hyena, Adam, and SGD using the same architecture and dataset. Results demonstrate that Hyena attained expedited convergence, reduced RMSE (0.022 ± 0.004 compared to 0.034 ± 0.005 for Adam and 0.036 ± 0.006 for SGD), and superior average accuracy (95.7% versus 92.7% and 91.5%, respectively).

Table 9. Price estimation

Crop	Actual Price Trend	Predicted Trend (HZEF)	Accuracy (%)	Confidence Interval (%)	description
Wheat	Higher than seasonal average	Higher	91	[87, 95]	Decision support only
Rice	Lower than seasonal average	Lower	89	[85, 93]	Market fluctuations considered
Maize	Higher than seasonal average	Higher	92	[88, 96]	Trend-based prediction
Soybean	Lower than seasonal average	Lower	90	[86, 94]	Relative price only
Barley	Higher than seasonal average	Higher	93	[89, 97]	Relative trend guidance

Table 10. User Study Evaluation of Voice Interface

Participant ID	Age	Farming Experience (Years)	Comprehension Rate (%)	Interaction Success (%)	Qualitative Feedback / Notes
P1	32	8	90	85	Understood growth stage guidance; needed repetition for price trends
P2	45	20	85	80	Comfortable with harvest timing prompts; dialect required slight adaptation
P3	27	5	95	90	Found interface intuitive; recommended adding confirmation prompts
P4	50	25	80	75	Some difficulty with unfamiliar terms; appreciated step-by-step instructions
P5	35	12	88	82	Clear instructions; suggested optional visual output for confirmation

Table 11. Ablation study

Optimizer	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)	RMSE (%)
Hyena	95.7 ± 0.9	95.3 ± 1.0	95.1 ± 1.2	95.2 ± 1.2	0.022 ± 0.004
Adam	92.7 ± 1.4	91.9 ± 1.6	91.2 ± 1.8	91.4 ± 1.5	0.034 ± 0.005
SGD	91.5 ± 1.6	90.8 ± 1.7	90.1 ± 1.9	90.3 ± 1.8	0.036 ± 0.006

These data indicate that Hyena Optimization enhances both predictive performance and stability, especially for multimodal fusion tasks, hence validating its preference over traditional optimizers. An ablation study was conducted to validate the selection of Hyena Optimization over conventional optimizers (Adam and SGD). Table 11 demonstrates that Hyena attains the maximum accuracy, precision, recall, and F-score, concurrently yielding the lowest RMSE. The enhanced convergence and gradient propagation in high-dimensional, multimodal feature spaces render Hyena especially useful for the HZEF framework, substantiating its superiority over traditional optimizers in this context.

Ethical and Practical Deployment Concerns

The proposed HZEF framework demonstrates robust predictive performance; however, its use in real-world scenarios must account for practical and ethical limitations. Internet connectivity and smartphone access may be constrained in rural agricultural areas, possibly limiting system availability. To address this, a lightweight

offline version of the model could be implemented on edge devices, requiring only periodic synchronization for updates. Voice assistant functionality may encounter difficulties due to regional dialect variations, ambient noise, or subpar microphones; implementing adaptive speech recognition models and noise filtering can enhance usability under these conditions. Moreover, energy efficiency is a crucial factor for field deployment; model inference has been refined to minimize computational and memory requirements to facilitate battery-powered operation. By recognizing these constraints and developing solutions for offline, low-power, and dialect-sensitive functionality, HZEF can serve as a pragmatic, ethically accountable decision-support tool for farmers without excessive reliance on idealized infrastructure.

Economic and Practical Validation

A historical backtesting examination of prior crop pricing data was done to see how useful the planned HZEF system would be in real life. The present study have contrasted the system's recommendations for relative price trends to other tactics, like harvesting on set calendar dates or using common farmer heuristics. The results indicate that a farmer adhering to HZEF instructions would have capitalised on a significant market designs, illustrating potential economic advantages that extend beyond conventional model accuracy measurements. A small-scale conceptual user study with representative farmers was done to test usability, even though a fully working voice interface hasn't been put into the practice. Participants were shown prototype outputs through a fake voice interface, and qualitative feedback was gathered on how well they understood, how clear the interaction was, and how valuable they thought it was. Initial feedback suggested that the interface concept is intuitive and actionable; nonetheless, variations in local dialects and levels of technical proficiency were recognised as possible obstacles. Deployment matters, such as working offline in locations with poor internet, supportive many languages, and making mobile devices use less power, are recognised as important areas for future effort. These factors will be examined in future research to ascertain the practical viability and accessibility of the system.

Future research

The Hyena ZFNet Encoder Framework or HZEF for short needs to be tested on datasets to see how well it really works. This means using pictures of plants from sources like PlantVillage and AgroNet to check if the Hyena ZFNet Encoder Framework can handle different types of crops, weather and cameras. If it can do this then it will be useful for more than the specific dataset we used. But there is a problem. We could not find any datasets that have all the information we need like pictures of crops, weather data and market prices from the same time. So we could not test the Hyena ZFNet Encoder Framework thoroughly as we wanted to. This is a limitation of our study. In the future we want to make the Hyena ZFNet Encoder Framework better by testing it on different datasets and using special techniques to help it work well in different situations. This will make the Hyena ZFNet Encoder Framework more reliable and useful in life. To see how well the Hyena ZFNet Encoder Framework really works we can try training it on one type of crop and then testing it on pictures of the crop from other places. This will show us if the Hyena ZFNet Encoder Framework can handle types of soil, weather and farming methods. For example we can train the Hyena ZFNet Encoder Framework on pictures of wheat. Then test it on wheat pictures from other datasets. This kind of testing is important because it helps us find out if the Hyena ZFNet Encoder Framework is working well or if it is just good at recognizing things it has seen before. Unfortunately we could not do this kind of testing well because we did not have enough datasets with the right information. We know it is something we need to do in the future to make the Hyena ZFNet Encoder Framework better. So to make the Hyena ZFNet Encoder Framework more useful we need to test it on different datasets and use special techniques to help it work well in different situations. This will make the Hyena ZFNet Encoder Framework more reliable and useful for farmers and other people who work with crops. The Hyena ZFNet Encoder Framework will be able to handle types of crops, weather and cameras and it will be a valuable tool, for people who need to recognize plants and understand how they are growing.

5. Conclusion

This study introduced the HZEF framework to forecast agricultural harvest timing and relative price trends by merging crop imagery, historical climatic data, and market price records. The data underwent preprocessing by

Gaussian filtering and histogram equalization, succeeded by feature extraction and hybrid multimodal fusion to produce predictions, incorporating voice output capabilities for effective decision assistance. The assessment using k-fold cross-validation and repeated trials indicates that HZEF achieves significant relative improvements over baseline methodologies, yielding around 3–5% higher accuracy and F-score, alongside a reduction of 35–40% in RMSE compared to Random Forest, XGBoost, LSTM, and Vision Transformer. Although these results demonstrate significant improvements, uncertainty arising from dataset size, market volatility, and illumination fluctuation is recognized, and all metrics are shown with confidence intervals. The framework serves as a pragmatic decision-support instrument, offering actionable insights on harvest timing and relative pricing patterns rather than precise forecasts. The present constraints encompass energy consumption and memory utilization, which were not examined in this study; subsequent research will concentrate on model optimization and lightweight deployment to improve practical applicability.

REFERENCES

1. S.D. Shingade, and R.P. Mudhalwadkar, *Analysis of crop prediction models using data analytics and ML techniques: a review*, Multimedia Tools and Applications, vol. 83, pp. 37813–37838, 2024.
2. M.V. Krishna, K. Swaroopa, and G. SwarnaLatha, *Crop yield prediction in India based on mayfly optimization empowered attention-bi-directional long short-term memory (LSTM)*, Multimedia Tools and Applications, vol. 83, pp. 29841–29858, 2024.
3. S. Boppudi, and J. S, *Deep ensemble model with hybrid intelligence technique for crop yield prediction*, Multimedia Tools and Applications, vol. 83, pp. 75709–75729, 2024.
4. R.R. Lamsal, P. Karthikeyan, P. Otero, and A. Ariza, *Design and implementation of Internet of Things (IoT) platform targeted for smallholder farmers: From Nepal perspective*, Agriculture, vol. 13, no. 10, pp. 1900, 2023.
5. D. Sinha, A.K. Maurya, G. Abdi, M. Majeed, R. Agarwal, R. Mukherjee, S. Ganguly, and R. Aziz, *Integrated genomic selection for accelerating breeding programs of climate-smart cereals*, Genes, vol. 14, no. 7, pp. 1484, 2023.
6. M.M. Kaya, Y. Taşkiran, A. Kanoğlu, A. Demirtaş, E. Zor, İ. Burçak, M.C. Nacak, and F.T. Akgul, *Designing a smart home management system with artificial intelligence and machine learning*, Technical Report, 2021.
7. S.I. Hassan, M.M. Alam, U. Illahi, M.A. Al Ghamdi, S.H. Almotiri, and M.M. Su'ud, *A systematic review on monitoring and advanced control strategies in smart agriculture*, IEEE Access, vol. 9, pp. 32517–32548, 2021.
8. S.C. Areed, S.A. Salloum, and K. Shaalan, *The role of knowledge management processes for enhancing and supporting innovative organizations: A systematic review*, in Recent Advances in Intelligent Systems and Smart Applications, Springer, pp. 143–161, 2020.
9. J. Steinke, J. Van Etten, A. Müller, B. Ortiz-Crespo, J. van de Gevel, S. Silvestri, and J. Priebe, *Tapping the full potential of the digital revolution for agricultural extension: An emerging innovation agenda*, International Journal of Agricultural Sustainability, vol. 19, no. 5–6, pp. 549–565, 2021.
10. P. Castillejo, G. Johansen, B. Cürüklü, S. Bilbao-Arechabala, B. Martínez-Rodríguez, L. Pomante, C. Rusu, J.F. Martínez-Ortega, C. Centofanti, M. Hakojärvi, M. Santic, and J. Häggman, *Aggregate farming in the cloud: The AFarCloud ECSEL project*, Microprocessors and Microsystems, vol. 78, pp. 103218, 2020.
11. E.M. Emeana, L. Trenchard, and K. Dehnen-Schmutz, *The revolution of mobile phone-enabled services for agricultural development (m-Agri services) in Africa: The challenges for sustainability*, Sustainability, vol. 12, no. 2, pp. 485, 2020.
12. G. Kakamoukas, P. Sarigiannidis, A. Maropoulos, T. Lagkas, K. Zaralis, and C. Karaiskou, *Towards climate smart farming—A reference architecture for integrated farming systems*, Telecom, vol. 2, no. 1, pp. 52–74, 2021.
13. M. Javaid, A. Haleem, R.P. Singh, and R. Suman, *Enhancing smart farming through the applications of Agriculture 4.0 technologies*, International Journal of Intelligent Networks, vol. 3, pp. 150–164, 2022.
14. K.R. Chowdhary, *Natural language processing*, in Fundamentals of Artificial Intelligence, Springer, pp. 603–649, 2020.
15. A. Bohr, and K. Memarzadeh, *The rise of artificial intelligence in healthcare applications*, in Artificial Intelligence in Healthcare, Academic Press, pp. 25–60, 2020.
16. D. Michelsanti, Z.H. Tan, S.X. Zhang, Y. Xu, M. Yu, and D. Yu, *An overview of deep-learning-based audio-visual speech enhancement and separation*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 1368–1396, 2021.
17. C. Zhang, Z. Yang, X. He, and L. Deng, *Multimodal intelligence: Representation learning, information fusion, and applications*, IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 3, pp. 478–493, 2020.
18. C.H.M.H. Saibaba, S.F. Waris, S.H. Raju, V. Sarma, V.C. Jadala, and C. Prasad, *Intelligent voice assistant by using OpenCV approach*, in Proceedings of the Second International Conference on Electronics and Sustainable Communication Systems, IEEE, 2021.
19. C.S.C. Dalim, M.S. Sunar, A. Dey, and M. Billinghurst, *Using augmented reality with speech input for non-native children's language learning*, International Journal of Human-Computer Studies, vol. 134, pp. 44–64, 2020.
20. E. Cambria, R. Mao, M. Chen, Z. Wang, and S.B. Ho, *Seven pillars for the future of artificial intelligence*, IEEE Intelligent Systems, vol. 38, no. 6, pp. 62–69, 2023.
21. D. Loukatos, A. Fragkos, and K.G. Arvanitis, *Exploiting voice recognition techniques to provide farm and greenhouse monitoring for elderly or disabled farmers*, in Bio-Economy and Agri-production, Academic Press, pp. 247–263, 2021.
22. D.M. Matilla, A.L. Murciego, D.M. Jiménez-Bravo, A.S. Mendes, and V.R.Q. Leithardt, *Low-cost edge computing devices and novel user interfaces for monitoring pivot irrigation systems based on IoT and LoRaWAN technologies*, Biosystems Engineering, vol. 223, pp. 14–29, 2022.
23. V. Nayak, and N.H. Sowmya, *Agroxpert—Farmer assistant*, Global Transitions Proceedings, vol. 2, no. 2, pp. 506–512, 2021.

24. A. Rahman, M. Xi, J.J. Dabrowski, J. McCulloch, S. Arnold, M. Rana, A. George, and M. Adcock, *An integrated framework of sensing, machine learning, and augmented reality for aquaculture prawn farm management*, Aquacultural Engineering, vol. 95, pp. 102192, 2021.
25. J. Mössinger, C. Troost, and T. Berger, *Bridging the gap between models and users: A lightweight mobile interface for optimized farming decisions*, Agricultural Systems, vol. 195, pp. 103315, 2022.
26. K.L. Raju, and V. Vijayaraghavan, *Architecture development with measurement index for agriculture decision-making system using IoT and machine learning*, Multimedia Tools and Applications, vol. 82, pp. 36119–36142, 2023.
27. A. Gupta, and P. Nahar, *Classification and yield prediction in smart agriculture system using IoT*, Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 8, pp. 10235–10244, 2023.
28. S. Si-Kaddour, L. Boubchir, and B. Daachi, *Deep face recognition based on an optimized deep neural network using ZFNet*, in Proceedings of the 20th ACS/IEEE International Conference on Computer Systems and Applications, IEEE, 2023.
29. P. Li, Y. Pei, and J. Li, *A comprehensive survey on design and application of autoencoder in deep learning*, Applied Soft Computing, vol. 138, pp. 110176, 2023.
30. L. Fan, J. Li, and J. Liu, *An enhanced spotted Hyena optimization algorithm and its application to engineering design scenario*, International Journal of Artificial Intelligence Tools, vol. 32, no. 6, pp. 2350019, 2023.
31. T. Yusifzada, *Evaluating the global impact of climate change on agricultural inflation: An innovative climate condition index approach*, Environment, Development and Sustainability, vol. 26, no. 7, pp. 18411–18438, 2024.
32. X. Cui, and Z. Zhong, *Climate change, cropland adjustments, and food security: Evidence from China*, Journal of Development Economics, vol. 167, pp. 103245, 2024.
33. E.A. Elsadek, K. Zhang, Y.A. Hamoud, A. Mousa, A. Awad, M. Abdallah, H. Shaghaleh, A.A.A. Hamad, M.T. Jamil, and A. Elbeltagi, *Impacts of climate change on rice yields in the Nile River Delta of Egypt*, Agricultural Water Management, vol. 292, pp. 108673, 2024.
34. A. Saleem, S. Anwar, T. Nawaz, S. Fahad, S. Saud, T. Ur Rahman, M.N.R. Khan, and T. Nawaz, *Securing a sustainable future: The climate change threat to agriculture, food security, and SDGs*, Journal of Umm Al-Qura University for Applied Sciences, pp. 1–17, 2024.
35. L. Hatton, R. Bañares-Alcántara, S. Sparrow, F. Lott, and N. Salmon, *Assessing the impact of climate change on the cost of production of green ammonia from offshore wind*, International Journal of Hydrogen Energy, vol. 49, pp. 635–643, 2024.
36. R.L. Manogna, V. Dharmaji, and S. Sarang, *A novel hybrid neural network based volatility forecasting of agricultural commodity prices*, Journal of Big Data, vol. 12, pp. 85, 2025.
37. Y. Min, Y.R. Kim, Y. Hyon, T. Ha, S. Lee, and J. Hyun, *RNN and GNN based prediction of agricultural prices with multivariate time series*, Scientific Reports, vol. 15, pp. 13681, 2025.
38. G.H.H. Nayak, et al., *Meta transformer: Leveraging metaheuristic algorithms for agricultural commodity price forecasting*, Journal of Big Data, vol. 12, pp. 13896, 2025.