DEEP FUSION WITH ATTENTION NEURAL NETWORK FOR SOIL HEALTH BASED CROP YIELD ESTIMATION USING REMOTE SENSING IMAGES

Vijaykumar P. Yele[1], R. R. Sedamkar[2], Sujata Alegavi[3]

[1]Department of Electronics & Telecommunication Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra-400101, India.
[2]Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra-400101, India.
[3]Department of Internet of Things, Thakur College of Engineering and Technology, Mumbai, Maharashtra-400101, India.

ABSTRACT

Crop yield estimation, vital for agricultural planning, incorporates weather, soil health, and technology. Utilizing remote sensing to analyze soil health enhances agricultural management and resource optimization. Despite challenges like data accuracy and cloud interference, the proposed multi-head cross attention with capsule energy valley network (mhca-cevn) tackles these issues. This research integrates sentinel-1 and sentinel-2 data with field measurements, employing advanced preprocessing and feature extraction methods, such as the guided multi-layer side window box filter and shearlet transform. The hybrid gold rush mantis search optimizer selects key features for a deep visual attention-based fusion method. The resulting mhca-cevn classification model achieves outstanding performance, with accuracy, sensitivity, error rate, f1-score, mean absolute percentage error, and symmetric mean absolute percentage error at 97.59%, 95.21%, 6.65%, 90.21%, 5.01%, and 0.042%, respectively. These metrics highlight the model's efficacy in addressing diverse crop yield challenges, establishing it as a robust solution for remote sensing.

KEYWORDS

Crop yield estimation, MHCA-CEVN, Guided multi-layer side window box filter and shearlet transform, Hybrid gold rush mantis search optimizer, Deep Visual Attention.

1. INTRODUCTION

Crop Yield Estimation (CYE) is crucial in modern agriculture, serving as a predictive tool to anticipate crop yields on specific parcels [1]. Essential for precision farming and resource management, accurate CYE integrates traditional methods, advanced technologies, and data analytics [2]. This process informs decision-making in resource allocation, risk mitigation, and overall agricultural planning. Evolution in CYE includes remote sensing, weather data analysis, and on-field sensors, providing comprehensive insights for optimizing farming practices and addressing challenges from changing environmental conditions [3].

David C. Wyld et al. (Eds): AISO, SIP, SOFTFM, NLDM, CRBL, BIOM, COMIT– 2024
pp. 01-22, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.141501

In CYE with Remote Sensing Images (RSI), satellite or aerial imagery efficiently evaluates agricultural productivity non-intrusively [4,5]. Remote sensing technologies capture spectral bands, extracting data on vegetation indices and crop health markers [6]. These images provide a panoramic view of large agricultural areas, enabling observation of spatial crop variations [7,8]. Advanced image processing, including spectral analysis and Machine Learning (ML), extracts meaningful insights. Correlating remote sensing metrics with ground truth data generates precise crop yield estimates, facilitating timely decision-making for farmers to optimize resource allocation and address challenges In the CYE domain, RSIare applied to assess soil health and predict crop yield. This innovative method utilizes satellite imagery and other distant sensing techniques to comprehensively evaluate soil and crop conditions [9,10]. Various remote sensing data, including microwave and optical data, monitor land-use changes, crop growth, soil moisture, salinity, and pest infestation levels. Analysing this diverse data aids in estimating crop yields, offering valuable insights for informed agricultural decision-making [11]. Incorporating ML-based approaches, proposed frameworks and models enhance understanding of soil health and crop dynamics, promoting sustainable and data-driven practices in agriculture [12].

Soil health-based CYE with RSI transforms sustainable agriculture. Using remote sensing, such as satellite imagery, offers farmers vital insights for precise management [13]. Optimizing CYE with a focus on soil health promotes resource efficiency, minimizing environmental impact. This empowers informed decisions, enhances productivity, and fosters environmentally responsible farming, contributing to resilient and adaptive agriculture for long-term global food production sustainability [14,15].

The major contributions of this suggested work are listed below:

- Presenting the enhanced guided multi-layer side window box filter for the preprocessing of initial satellite imagery, this innovative technique is employed to refine and optimize the quality of the input images.
- For feature extraction, the application of an enhanced non subsampled shearlet transform is employed to achieve superior precision and representation.
- The numerical data undergoes preprocessing through data normalization and transformation techniques. Selection of essential features is performed using the hybrid Gold Rush Mantis Search (GRMS) optimizer.
- The integration of two features is achieved through the utilization of a deep visual attention-based fusion method. Subsequently, the combined features are input into the MHCA-CEVN for CYE.

The next parts of this research work are set up in the manner described: section 2 looks into recent strategies relevant to this work, section 3 explains the recommended design, section 4 details the outcomes and analysis, while section 5 summarizes and concludes this work.

2. LITERATURE SURVEY

The current body of literature delves into a variety of techniques and aspects, as evidenced by references [16-25]. This extensive exploration encompasses diverse methodologies and considerations derived from the cited sources, contributing to a comprehensive understanding of the subject matter.

In 2022, Tripathi et al. [16] introduced Deep Learning Multi-Layer Perceptron (DLMLP) neural networks, diverging from optical data-centred approaches to forecast wheat crop yield by estimating soil health parameters. The study emphasized the pivotal role of healthy soil, affirming its significant impact on achieving optimal crop yields. In 2022, Seydi et al. [17] introduced an

innovative Deep Learning (DL) method for mapping seven crop and three non-crop classes. Employing time-series Normalized Difference Vegetation Index (NDVI) for its dynamic crop representation, the model outperformed sophisticated supervised learning methods in both statistical and visual assessments, underscoring its superior effectiveness among advanced classification techniques. In 2022, Ali et al. [18] introduced an innovative approach called Supervised Nearest Neighbourhood Object-based Classification (SNNOC) for assessing the geographical spread of rice cultivation and production. The yield prediction involved utilizing an empirical model based on NDVI and Leaf Area Index (LAI) for accurate estimations. In 2022, Ma et al. [19] introduced the Shuffled complex evolution with Principal component analysis-University of California, Irvine (SP-UCI) optimization algorithm, applying it to assimilate remote sensing data into a simple yield estimation and crop growth model. The assimilation system, evaluated for reliability and accuracy using two years of winter wheat growth data under varying conditions, consistently aligned with actual measurements for leaf area index, biomass, and yield.

Table 1: Assessing various existing methods, reviewing their strengths and weaknesses

AUTHORS	METHODS	ADVANTAGES	DISADVANTAGES
Tripathi et al. [16]	DLMLP	• High precision. • Better performance.	• Expanding locations improves evaluation thoroughness.
Seydi et al. [17]	DL	• Exceptional precision achieved across diverse NDVI datasets.	• Framework effectiveness varies across diverse regions. • Optimizing framework may need extensive experimentation.
Ali et al. [18]	SNNOC	• Yields high accuracy and kappa coefficient. • Provides timely insights before harvest.	• Demands resources, expertise for model. • Deviations from assumptions affect SEBAL model prediction reliability.
Ma et al. [19]	SP-UCI	• Improved effectiveness.	• Increase data types for better representation.
Meraj et al. [20]	CASA	• Strong correlation confirms approach's reliability, accuracy. • Method holds promise for regional crop yield.	• Parameter inaccuracies may bias yield estimates. • Temporal scope limits model's interannual variability.
Ma et al. [21]	AMM-FuseNet	• Excels with minimal training data. • Minimal accuracy reduction.	• Dataset specificity limits generalizability to diverse regions.
Uribeetxebarria et al. [22]	CatBoost	• Improves wheat yield prediction accuracy.	• Algorithm needs rare high-resolution yield data.
Yeasin et al. [23]	ML	• Precise spatial monitoring with high resolution.	• Validate over broader area, larger sample is needed.
Trivedi et al. [24]	CMT	• Improved adaptability.	• Optimize sampling for model improvement.
Singha et al. [25]	CC	• Sentinel-1 monitors tuber quality and yield.	• Ignores future data access changes.
Munaganuri et al. [35]	DDQGCN	• Reduced response delays and enhanced precision	• Exploring complex neural networks further enhance predictive capabilities.
Shao et al. [36]	ML	• High accuracy in estimating crop water consumption.	• No significant benefit from additional data types beyond multispectral.

Yang et al. [37]	ML	• Enhanced generalization capabilities of the model.	• increased computational complexity and require extensive tuning.
Ramzan et al. [38]	DNN	• Optimal forecasting accuracy	• Limited predictor scope and data scarcity hinders model accuracy.

In 2022, Meraj et al. [20] introduced a ML algorithm for wheat acreage estimation. Various ML classification algorithms were employed, and the Carnegie-Ames-Stanford Approach (CASA) system was applied to calculate wheat yield. Wheat acreage was determined through supervised classification, utilizing Random Forest (RF) and Support Vector Machine (SVM) classifiers. The classification accuracy of these methods was then compared based on ground-truthing. In 2022, Ma et al. [21] presented the Attention-based Multi-Modal image Fusion Network (AMM-FuseNet). The network leveraged a hybrid strategy involving the channel attention mechanism and Densely connected Atrous Spatial Pyramid Pooling (DenseASPP). The performance of the proposed network demonstrated competitiveness against leading-edge segmentation networks like DeepLabv3+, SegNet, and Unet. In 2022, Uribeetxebarria et al. [22] optimized wheat yield prediction using sentinel-1, sentinel-2 data with the Categorical Boosting (CatBoost) algorithm. They evaluated the approach on 39 fields, suggesting future studies consider high-resolution meteorological and edaphic variables for a more comprehensive analysis. In 2022, Yeasin et al. [23] applied ML techniques to assess sugarcane phenology. The study explored individual and combined data effectiveness, evaluated temporal behaviors, and suggested broader-scale validation studies with a larger sample size.

In 2023, Trivedi et al. [24] advocated an efficient Cropland Mapping Technique (CMT) with focus on sentinel-2 and sentinel-1 metrics. Elevation and texture-based features played crucial roles, influencing crop distribution and spatial differences in small, irregular fields. The study advised image stratification and optimized sampling for enhanced model performance and transferability. In 2022, Singha et al. [25] demonstrated an efficient approach for monitoring potato cropping areas using polarized images from sentinel-1 C band in different phenological stages. The google earth engine Cloud Computing (CC) framework allowed rapid pre-processing, facilitating quick data acquisition and processing. Table 1 assessed existing methods with their benefits and limitations.

In 2024, Munaganuri et al. [35] introduced the Deep Dyna Q Graph Convolutional Network (DDQGCN), which classified selected features into irrigation schedule classes. This network enhanced classification efficiency across diverse image sources and improved adaptability to different crop types and environmental conditions.In 2023, Shao et al. [36] developed models for estimating crop coefficients for irrigated maize in Northwest China's semi-arid region using Unmanned Aerial Vehicle (UAV) remote sensing and ML techniques. Six ML algorithms and deep neural network were employed for model development.

In 2024, Yang et al. [37] explored UAV multi-sensor data fusion and ML for wheat yield prediction. They proposed layered ensemble method, feature-prioritized, and basic averaging ensemble techniques, which improved prediction accuracy.In 2023, Ramzan et al. [38] combined agrometeorological and remote sensing data to predict farm-level tea yield. The study proposed a Deep Neural Network (DNN) with three hidden layers, optimized through Bayesian search, for complex non-linear modelling using Landsat imagery.

Problem statement

Assessing soil health is crucial as it serves as a vital natural resource supporting all life on earth and directly impacting crop yields. Global studies use remote sensing for agricultural monitoring and yield estimation, but challenges arise in addressing crop variability, growth stages, and

environmental conditions. Agriculture productivity ties closely to climatic elements, fertilizer use, and soil fertility. Remote sensing faces difficulties in discriminating crops and dealing with cloud cover. Model calibration, validation, and multi sensor integration add complexity. This study proposes a solution employing a deep fusion with an attention neural network for accurate CYE based on soil health using RSI, addressing these challenges.

3. PROPOSED METHODOLOGY FOR CROP YIELD ESTIMATION USING MHCA-CEVN

The process begins with the utilization of remotely sensed data from Sentinel-1 Microwave satellites and Sentinel-2 optical data, complemented by field data, to assess key soil health parameters: Soil Moisture, Soil Salinity, and Soil Organic Carbon (SOC). The derived soil health parameters, along with the optical remote sensing data, are then used to estimate crop yield. The research employs two primary datasets: remotely sensed images and numerical data, including soil health parameters, area, rainfall, temperature, air quality index, and quantities of SO2 and NO2.

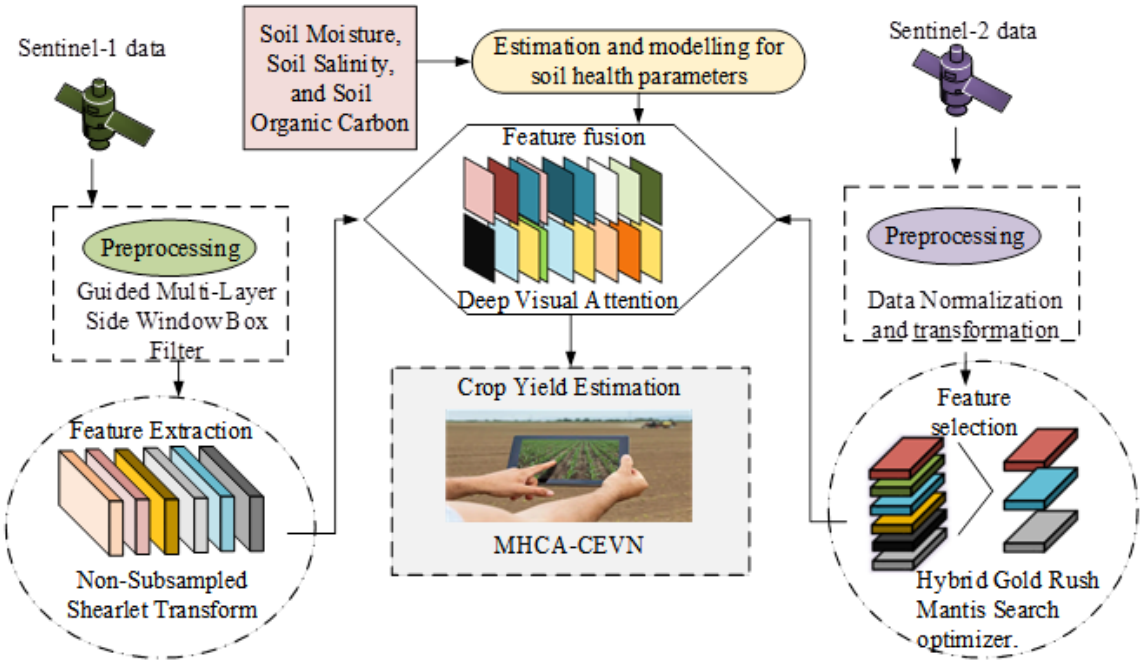

Figure 1. System model

The process starts with the pre-processing of satellite images using the enhanced guided multi-layer side window box filter and feature extraction via an improved non-subsampled shearlet transform. numerical data undergo normalization and transformation, with crucial features selected through the hybrid gold rush mantis search optimizer. A deep visual attention-based fusion method integrates the two feature sets, which are subsequently fed into a multi-head cross-attention network combined with a capsule energy valley network for crop yield estimation. The performance of the proposed method is evaluated using various metrics. Ultimately, the approach aims to enhance the precision of crop yield predictions, offering potential improvements in farming techniques to mitigate crop losses. Figure 1 shows the proposed system model.

3.1. Enhanced Guided Multi-Layer Side Window Box Filter for Preprocessing the Satellite Image

Satellite image preprocessing is crucial for enhancing quality and ensuring accurate analysis by addressing inherent noise, correcting atmospheric interferences, and standardizing radiometric values. The enhanced guided multi-layer side window box filter is a sophisticated technique for preprocessing satellite images. By addressing noise, atmospheric effects, and radiometric inconsistencies, it improves image quality and enables more accurate analysis. The filter's multi-layer and guided mechanisms, combined with the side window approach, make it a versatile tool for various image processing tasks.The guided filter improves image quality while preserving edges, contributing to optimized satellite image preprocessing for superior clarity in subsequent analytical processes [26]. Commencing the process, the input satellite images serve as the initial input for preprocessing, employing the enhanced guided multi-layer side window box filter.

The filtered equation is represented by Equation (1):

$$P_{SWF}(a,b)=\sum_{k=1}^{r}\sum_{l=1}^{t} w(k,l).P_{org}(a-k,b-l) \tag{1}$$

where, $P_{SWF}(a,b)$ indicates the filtered images, $P_{org}(a,b)$ is the original image, $w(k,l)$is the window function. The original satellite image is fed into the filter. The filter applies the guided multi-layer side window box approach to refine the image. The filtered image is obtained, which serves as the input for the subsequent feature extraction process.

In the realm of side window filtering, aligning the window's side or corner with the edge of an object in the image enhances adaptability and performance in a range of image fusion tasks. The filtered image, obtained as the output of the pre-processing, serves as the input for the subsequent feature extraction process.

3.2. Improved Non Subsampled Shearlet Transform for Feature Extraction

After preprocessing satellite images, feature extraction becomes crucial, distilling essential information for precision. It identifies patterns, reduces dimensionality, enhancing interpretability for tasks like classification and object recognition.

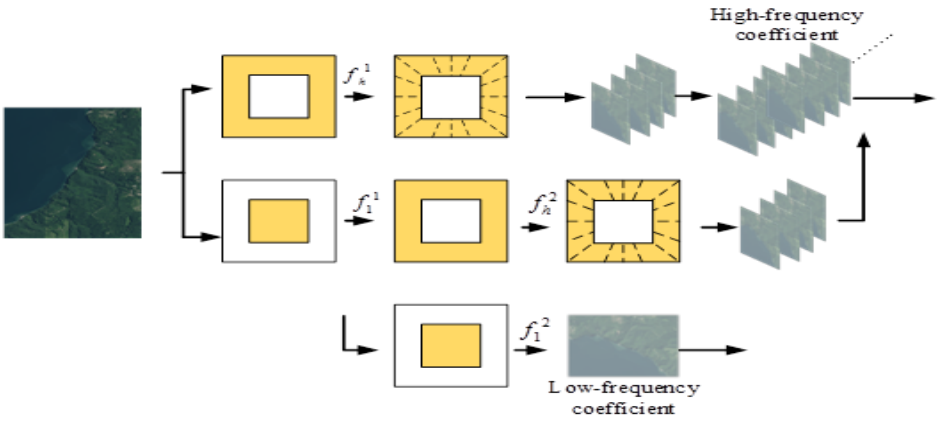

Figure 2. Block diagram of improved non subsampled shearlet transform.

The selection of the transform for feature extraction from pre-processed satellite images stems from its unique ability to capture intricate details, edges, and textures at various scales and orientations. This adaptability makes it a powerful tool, enhancing the capability to discern and analyse complex spatial patterns in satellite imagery [27].

Figure 2 shows the block diagram of improved non subsampled shearlet transform. f_1 is the frequency of the wavelet extracting information from signals. f_k represents the central frequency of the k^{th} sub band during signal decomposition into frequency ranges.

The enhanced non-subsampled shearlet transform enhances feature extraction by first decomposing signals into frequency sub-bands using wavelet transforms. It then applies shearlet filters at various scales and orientations to capture directional features and textures. The process involves using scale and shear matrices to control the filter application, yielding coefficients that represent the signal's characteristics. These coefficients are combined to form a comprehensive feature vector, capturing essential image information across multiple scales and directions. This method improves edge and texture detection while ensuring robustness and shift-invariance, making it highly effective for complex image analysis tasks.

Introduce the scale matrix M_s, where s denotes the scale parameters of the image, calculated by Equation (2). Following the decomposition of the sub band image using the shear matrix in the direction of non-subsampled shearlet, the quantity v represents the number of shearlet filters generated at a specific scale s. Equation (3) and Equation (4) present the formula for image redundancy V_s following transformer decomposition iterations within a total scale of S.

$$M_s = \begin{bmatrix} 2^s & 0 \\ 0 & 2^{\frac{s}{2}} \end{bmatrix} (s \in Z) \tag{2}$$

$$v = 2 \times \left(2^s - 1\right) + 2 = 2^{s+1} \tag{3}$$

$$V_s = \sum_{s=1}^{S} \left(2^{s+1}\right) + 1 \tag{4}$$

The formulation of the feature extraction is expressed by Equation (5).

$$W_f(a,b,\theta) = \sum_{k=1}^{r} \sum_{l=1}^{t} s_{k,l} \,.\, \Psi_{k,l}(a,b) \tag{5}$$

where, $s_{k,l}$ is the coefficients of the shearlet basis functions, $\Psi_{k,l}(a,b)$ are the shearlet basis functions at scale k and direction l, and $W_f(a,b,\theta)$ represents the improved non subsampled shearlet transform coefficient.

Table 2. Improved non subsampled shearlet transform algorithm

Algorithm: improved non subsampled Shearlet transform

Input: Pre-processed image $P_{SIFT}(a,b)$, total scale $S \in n$

1. Transform the input image P from spatial domain to frequency domain using a 2D Fast Fourier Transformer.
2. Generate shearlet spectrum $\Psi_{k,j}(a,b)$ for a given shape and scale.
3. **for** $s = 1,\ldots,S$ **do**
4. **for** $v_s = 1,\ldots,x$ **do**
5. Calculate shearlet coefficient $W_f(a,b,\theta)$ obtained by convolving frequency image with shearlet spectrum $\Psi_{k,j}(a,b)$
6. Store the coefficients $W_f(a,b,\theta)$ in the output array.
7. **end for**
8. **end for**
9. Separating $W_f(a,b,\theta)$ into low frequency and high frequency parts and applying the data weighting method on high frequency part to obtain $W_f(a,b,\theta)$

Output: The frequency coefficients $W_f(a,b,\theta)$

Table 2 provides the improved non subsampledshearlet transform algorithm. The image processing workflow begins by converting the input image into the frequency domain using a 2 dimension Fast Fourier Transform (FFT), facilitating subsequent analysis by simplifying the image representation. A shearlet spectrum is then generated for each desired scale and orientation, providing a structured framework to characterize shearlet properties for analysing specific image features. Shearlet coefficients are computed by convolving the frequency domain image with the corresponding shearlet spectrum, offering a detailed representation of the image content at various scales and orientations. These calculated coefficients are stored for efficient access in subsequent processing. The frequencies are split into low and high groups, where low-frequency coefficients capture overall structure, and high-frequency coefficients highlight intricate details. A data weighting method is applied to high-frequency coefficients to reduce noise and enhance sparsity, leading to a streamlined representation. Finally, essential features are extracted from high-magnitude shearlet coefficients, emphasizing their significance in capturing distinctive aspects of the image and improving the accuracy of feature extraction for further analysis.

3.3. Data Normalization and Transformation Technique for Preprocessing of Numerical Data

Data normalization and transformation techniques are essential preprocessing steps for numerical data. Normalization encompasses bringing the data to a standardized range, often between 0 and 1, ensuring consistent comparisons. Transformation modifies the distribution of data, addressing issues like skewness. Common methods involve min-max scaling, Z-score normalization, and logarithmic or power transformations. These techniques enhance model performance, mitigate the impact of outliers, and facilitate meaningful interpretation of numerical features in various data analysis and ML applications.

The formula for min-max normalization technique for data normalization and transformation, is given by Equation (6)

$$r' = \frac{r - r_{\min}}{r_{\max} - r_{\min}} \tag{6}$$

where, r is the original image, $r_{\min}$ and $r_{\max}$ are the minimum and maximum values of the characteristics respectively and r' is the output image. This equation standardizes feature values, transforming them to a range from 0 to 1, facilitating effective comparison and analysis.

3.4. Hybrid Gold Rush Mantis Search Optimizer for Feature Selection

Essential for streamlined data analysis and ML, feature selection trims irrelevant or redundant features in large datasets, curbing overfitting and computational complexity. By choosing the most relevant subset, it enhances model efficiency, interpretability, and accuracy, fostering a deeper understanding of underlying data patterns and aiding effective decision-making across diverse domains.

❖ Mantis Search optimization algorithm

Mantis Search Optimization (MSA) algorithm is motivated by the specific predatory conduct and occasional occurrences of sexual cannibalism observed in praying mantises, offering an innovative approach to optimization, MSA is represented by Equation (7).

$$A_c^{t+1} = Y_c^t + r_n \times \left(Y_c - Y_a\right) \tag{7}$$

A_c^{t+1} is updated value for parameter at time t. Y_c^t is current value of Y_c at time t. Y_c is variable representing characteristic. Y_a is variable representing a different characteristic and r_n is scalar or coefficient used in the update process.

❖ Gold Rush optimization algorithm

The advantage of this is the selection of different movement options without a force in decision-making. So, this optimization is used in proposed method for fusion of input features. The representation of gold rush optimizationis given by Equation (8)

$$B_1(c) = \sum_{c=1}^{n} Y_c^2 \tag{8}$$

where, $B_1(c)$ is represents fusion parameter using optimization algorithm.

❖ Hybridization of Gold Rush Mantis Search optimization algorithm

The process involves combining the GRMS optimization algorithm with hybrid techniques to perform feature selection on input data. This hybrid approach aims to boost the efficiency and effectiveness of the attribute selection process, optimizing the selection of relevant features for subsequent analysis or modelling. The implementation of GRMS for feature selection treats each feature as an independent variable. In this model, an individual represents a subset of features, and the fitness function evaluates the performance of a model built with those selected features [28-30].

The hybridization of mantis search optimization algorithm and gold rush optimization algorithm gives the Equation (9) and denoted as $Y(q)$.

$$Y(q) = F\left(A_c(q) + B_c(q)\right) \tag{9}$$

where, $A_c(q)$is output of MSA algorithm and $B_c(q)$is output of GRMS algorithm. Fitness evaluation equation formulated by Equation (10).

$$F_t = \sum_{c}^{n} Y_c . Score\left(Y(q)\right) \tag{10}$$

where F_t is fitness function. The final output of feature selection is represented by Equation (11)

$$r_S = \sum_{c}^{n} M\left(r^{'}, Y, z, N\right) \tag{11}$$

where, r_S is the selection output, $r^{'}$ is the input image, Y is the target variable, z size of features, N number of features to be selected.

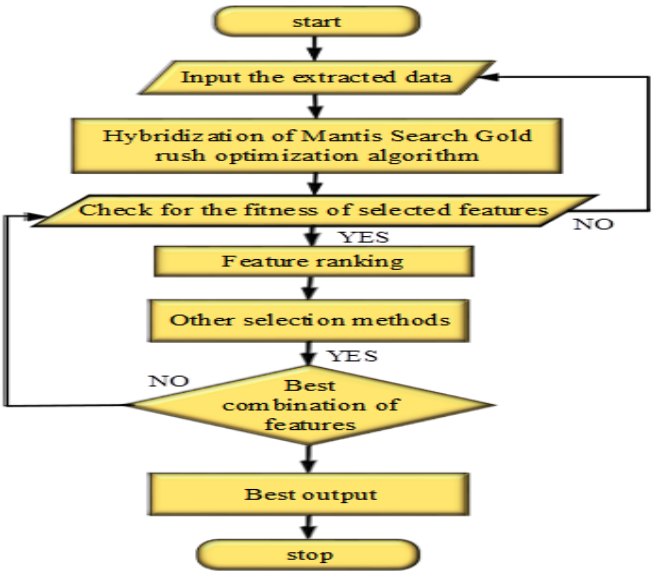

Figure 3. Flow chart of GRMS.

The below Figure 3. shows the flowchart for GRMS algorithm. The process begins with step 1, where pre-processed data is inputted into the hybrid algorithm, encompassing dataset features or any pertinent information essential for its operation. Subsequently, in step 2, the gold rush and mantis search optimization algorithm undergo hybridization, leveraging the strengths of both to craft a more potent and efficient algorithm. Moving to step 3, the fitness of selected features is assessed using Equation (10), with those exhibiting low fitness discarded and high-fitness features retained. Step 4 involves feature ranking based on their fitness, establishing a hierarchical order for the next phase. In step 5, various selection methods beyond feature ranking are explored to identify the ideal combination of features. Step 6 culminates in the prime feature combination identification, guided by the ranked features and selection methods, aiming for optimal performance on the validation dataset. Finally, step 7 involves training the hybrid algorithm on the training dataset using the chosen features, ultimately generating the best output on the test dataset. This comprehensive process underscores the systematic approach to enhancing algorithmic performance through hybridization and feature selection.

3.5. Deepvisual Attention for Feature Fusion

Feature fusion facilitates the amalgamation of varied attributes, fostering a comprehensive comprehension of the agricultural landscape. The integration of information from disparate sources, including satellite imagery, weather data, and soil quality assessments, augments the model's capacity to grasp the synergies and interdependencies among variables. This, in turn, results in a more precise and all-encompassing estimation of crop yield by capturing the intricate relationships within the agricultural ecosystem [31].

In remote sensing feature fusion, the deep attention network employs for fusion. The process begins with an encoder that extracts features from various data types, including NDVI, hyperspectral images, moisture images, building/road data, crop land information, and environmental data (rainfall, temperature, humidity). The architecture integrates attention mechanisms that assign weights based on feature significance, focusing on relevant spatial or spectral regions during fusion. The decoder then reconstructs the fused features, generating a comprehensive representation that enhances the model's ability to identify elaborate designs and interrelations within the remote sensing data.
During this procedure, the two features derived from the previously mentioned step specifically, the images extracted through non subsampled shearlet transform and the selected features obtained through hybrid gold rush mantis search optimize are combined through the application of deep visual attention.

Feature fusion equation is represented by Equation (12)

$$F = \sum_{c=1}^{N} \alpha_c \otimes P_c \tag{12}$$

where, F represents the fused output N is number of features to be fused, P_c is c^{th} features to be fused and α_c attention weights corresponding to the c^{th} feature. $\otimes$ represents element-wise multiplication, and $\sum$ denotes the summation over all input features.

3.6. Multi-Head Cross Attention Network with Capsule Energy Valley Network Based Classification Method

The integration of a MHCA-CEVN holds promise for advancing CYE. Multi head cross attention proves pivotal by allowing selective focus on pertinent information from diverse sources, such as remote sensing and weather data. This approach ensures more accurate predictions by considering relevant information across modalities and learning complex relationships between different data sources.

The input features are initially processed through a 1×1 convolution layer to reduce the number of channels. Following this, 3×3, 1×3, and 3×1 convolution kernels are employed to accurately map spatial interactions. The spatial attention unit comprises a set of four convolution layers paired with one activation function, designed to capture local features at multiple scales. The channel attention block includes two linear layers along with one activation function. Cross-head attention, which combines spatial and channel attention units while maintaining their independence, is also utilized. The above process is mathematically expressed by Equation (13).

$$Spatial_i = \vec{W} \times head_i\left(\eta_{spatial}, \vec{W}\right), i \in \{1, m\}$$

$$Channel_i = Spatial_i \times head_i^{'}\left(\eta_{channel}, Spatial_i\right), i \in \{1, m\} \tag{13}$$

where, m indicates the number of cross attention mechanisms, $\vec{W}$ isthe weighted featuremap, $head_i$ and $head_i^{'}$ are referred to as the spatial attention component and the channel attention component respectively, $\eta_{spatual}$ and $\eta_{channel}$ represent their coefficients, $Spatial_i$ and $Channel_i$ signify the outputs of the i^{th} spatial attention and channel attention, individually.

Meanwhile, capsule networks, with their unique spatial relationship and part whole hierarchy capturing ability, enhance the analysis of complex features within images, crucial for accurate yield predictions. Integrating these methodologies within the energy valley network framework could revolutionize CYE, contributing to improved accuracy, resource allocation, and agricultural sustainability [32-34].

The predictions for the next layer capsules are determined through a multiplication with a matrix of weights, as illustrated by Equation (14).

$$Cap_i = B_i\, X \tag{14}$$

where, Cap_i is input of capsule i in present layer, B_i is the weights matrix. X represents the input data (such as remote sensing imagery, meteorological data, and soil information)

$$d_i = \frac{\|Cap_i\|^2}{1+\|Cap_i\|^2}\frac{Cap_i}{\|Cap_i\|} \tag{15}$$

This Equation (15) ensures that the length of the capsule output vector d_i represents the probability. It non-linearly scales the input vector Cap_i to have a length between 0 and 1, preserving the direction of Cap_i. Energy valley network defined by Equation (16).

$$En(Out) = \sum_i \left\| Out_i - d_i \right\|^2 \tag{16}$$

The energy function $En(Out)$ measures the discrepancy between the capsule output Out_i and the vector d_i. The goal is to minimize this energy, which aligns the capsule outputs with the normalized inputs. Out_i is calculated by $Out_i = \sum_i R_{ij}\, d_i$, in that R_{ij} is routing coefficients.

The formula for estimating crop yield is expressed by Equation (17).

$$E_Y = f\left(M\left(Spatial_i, channel_i\right), E\left(Out_i\right)\right) \tag{17}$$

where, the input data includes information acquired from remote sensing platforms (such as satellite or aerial imagery), meteorological information (encompassing factors like rainfall, temperature, and humidity), soil information (including details such as soil type, nutrient content, and water availability), and past crop production data specific to the field or crop in question are extracted, $M(Spatial_i, channel_i)$ is result generated by the multi-head cross-attention network, acquiring insights into the connections among features extracted from diverse data sources $E(Out_i)$ is result obtained from the capsule energy valley network, acquiring hierarchical representations of both image data and historical yield data, f is a function that integrates the outcomes of both network methods to produce the ultimate yield estimation. Figure 4 shows the block diagram of MHCA-CEVN based classification method.

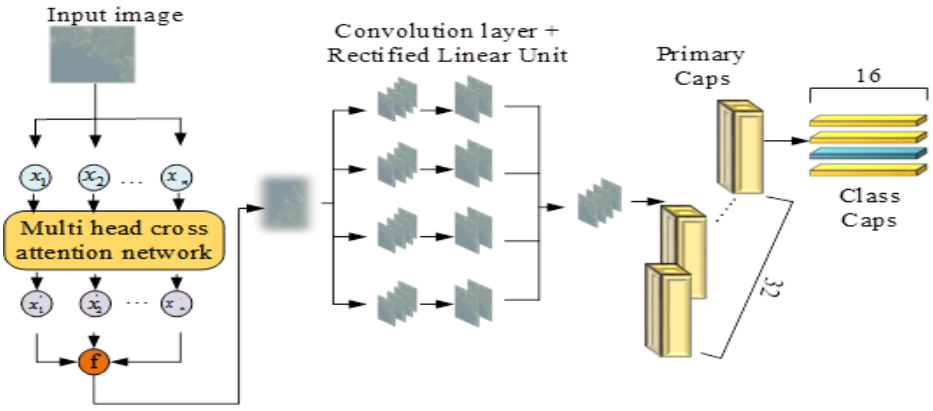

Figure 4. Block diagram of MHCA-CEVN based classification method

The MHCA-CEVN methodology integrates advanced techniques, including enhanced satellite image preprocessing, non-subsampled shearlet-based feature extraction, and hybrid GRMS for crucial feature selection. The amalgamated features are input into MHCA-CEVN, a multi-head cross attention network with capsule energy valley network, enhancing crop yield estimation accuracy through selective focus on diverse data sources and improved comprehension of intricate agricultural patterns, aiming to revolutionize accuracy, resource allocation, and sustainability in agriculture

4. RESULTS

In this section, various evaluation criteria, including Symmetric Mean Absolute Percentage Error (SMAPE), accuracy, sensitivity, error rate, F1-Score, and Mean Absolute Percentage Error (MAPE) functions, are examined. These parameters offer a comprehensive evaluation of the accuracy and dependability of identified changes, aiding in assessing the efficacy of the proposed technique.

4.1. Experimental Setup

The study utilized spyder anaconda navigator on windows 10 with python 3.10, operating on a robust system featuring 32 GB of RAM and an Intel(R) Core (TM) GeForce RTX 1080 super GPU @ 3.40GHz processor. After finalizing the network design, a comprehensive analysis explored factors like epoch count, repetitions, and a fixed learning rate of 0.001. With a set iteration value of 100, this detailed approach aimed to reveal variable interactions and their impact on experimental outcomes.

4.2. Dataset Description

The study employed Synthetic Aperture RADAR (SAR) and optical data from European Space Agency (ESA)sentinel-1 and sentinel-2 satellites as the primary remote sensing dataset. On-site field data were gathered using a Spectrum Technologies Field Scout SMEC 300 soil sensor. Samples of soil from the field were dispatched to Soil Organic Carbon (SOC) in %and pH determination. Crop yield information for the seasons 2018–19 to 2022–23 was obtained from Rupnagar's farming community [16]. Detailed dataset information is given in Table 3. Figures 5 (a-e) represents geospatial overview and wheat crop analysis in Rupnagar district, Punjab.

Table 3: Dataset used

Data	**Category**	**Resolution of space**	**Date of obtaining**
Sentil-1	SAR/Microwave channels with polarizations channels.	5 m × 20 m, multi looked to 14 m	21/11/2018, 23/12/2018, 24/01/2019, 20/11/2019, 23/12/2019, 21/01/2020, 24/11/2020, 23/12/2020, 25/01/2021, 20/11/2021, 19/12/2021, 21/01/2022, 21/11/2022, 28/12/2022, 21/01/2023.
Sentinel-2 A, 2B	Optical/multispectral sentinel-2's Band 3 is Near Infrared (NIR), and Band 8 is Short Wave Infrared (SWIR). Band 4 represents red, and Band 8 represents NIR. To account for varying spatial resolutions in sentinel-2's spectral bands, all 13 bands were resampled to a uniform 14 m spatial resolution.	10 m, modified to 14 m	21/11/2018, 23/12/2018, 24/01/2019, 20/11/2019, 23/12/2019, 21/01/2020, 24/11/2020, 23/12/2020, 25/01/2021, 20/11/2021, 19/12/2021, 21/01/2022, 21/11/2022, 28/12/2022, 21/01/2023
Field Data	Conductance (mS/m), Soil Moisture Percentage (%), and Temperature (^0C)	-	21/11/2022, 22/12/2022, 21/01/2023
Laboratory test results.	SOC (in %) and ph.	-	21/11/2022, 22/12/2022, 21/01/2023.
Wheat yield statistics.	Collected from Rupnagar farming survey, supplemented by government portal- https://aps.dac.gov.in/APY/Public_Report1.aspx	-	2018–19, 2019–20, 2020–21, 2021–22, 2022–23 Seasons.

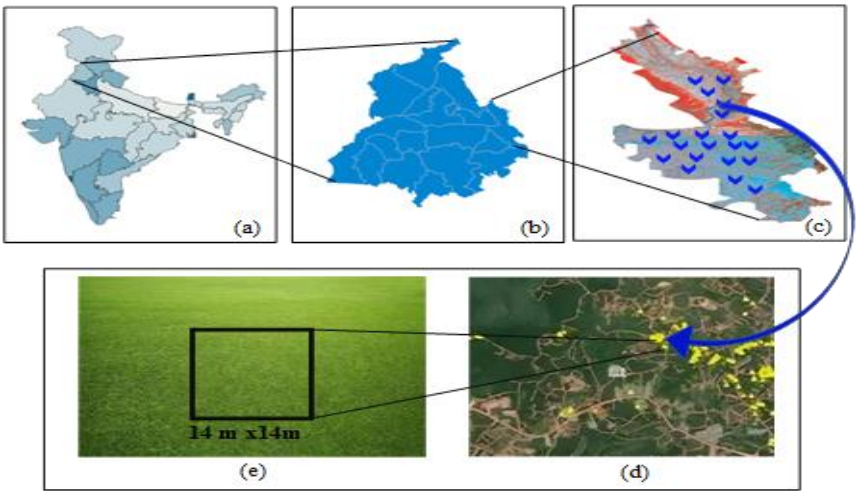

Figure 5. (a) Map of Indian states with Punjab highlighted, (b) Punjab state pinpointing Rupnagar district in red, (c) Typical incorrect colour composite of Rupnagar district, (d) Plantation areas in the study area cultivating wheat excluding other land use/land cover features, and (e) Wheat crop field displaying a pixel area designated for on-site data accumulation.

4.3. Performance Metrics

The evaluation criteria, which include accuracy, sensitivity, error rate, F1-score, MAPE, and SMAPE, are analysed.

Table 4 enumerates the equations used to derive various parameters. P stands for precision and R stands for recall. True Positives (TP) indicate the count of accurate positive predictions, True Negatives (TN) represent the count of accurate negative predictions, False Positives (FP) denote the count of incorrect positive predictions, False Negatives (FN) signify the count of incorrect negative predictions. A stand for accuracy, S represents recall, F_1 is the F1-score, and ER denoted error rate. O_k is the original value for observation k, P_k is predicted value for observation k, n is the number of observations. y_i is the actual value of the dependent variable (actual crop yield) for the i^{th} observation $\hat{y}_i$ is the predicted value of the dependent variable (predicted crop yield) for the i^{th} observation. $\bar{y}$ is the mean of the actual values of the dependent variable.

Table 4. Formulas used

Parameter used	**Formula**
Accuracy	$A = \frac{TP+TN}{TP+FN+FN+FP}$
F1-score	$F_1 = \frac{2PR}{P+R}$
Error Rate	$ER = \frac{No.Of\ Incorrect\ \ predictions}{TotalNo.Of\ predictions} \times 100$
MAPE	$MAPE = \frac{1}{n}\sum_{k=1}^{n}\left\lvert\frac{O_k - P_k}{O_k}\right\rvert \times 100$

SMAPE	$SMAPE = \frac{1}{n}\sum_{k=1}^{n}\left\|\frac{O_k - P_k}{\frac{\|O_k\| + \|P_k\|}{2}}\right\| \times 100$
Sensitivity	$S = \frac{TP}{TP + FN}$
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)}$

4.4. Comparative Assessment

The experimental investigation of CYE using the proposed technique includes the examination of accuracy, sensitivity, error rate, F1-score, MAPE, and SMAPE. These metrics are computed and juxtaposed with traditional methods.

Figure 6 gives accuracy investigation for CYE in RSI, evaluating precision by identifying changes and unchanged areas relative to total pixels or regions. The presented approach shows superior accuracy when contrasted to DLMLP [16], SNNOC [18], CASA [20], and ML [23], achieving 38.53%, 2.72%, 14.08%, and 10.89% greater accuracy, respectively, at 100 epochs.Sensitivity analysis, shown in Figure 7, indicates the proposed approach surpasses ML [23], CMT [24], and CC [25], achieving improvements of 7.80%, 22.06%, and 3.87% at 100 epochs, respectively.

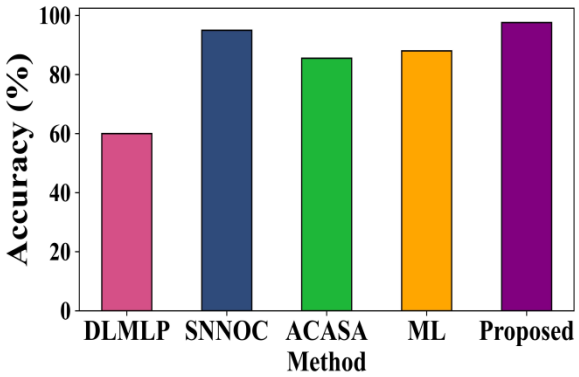

Figure 6. accuracy.

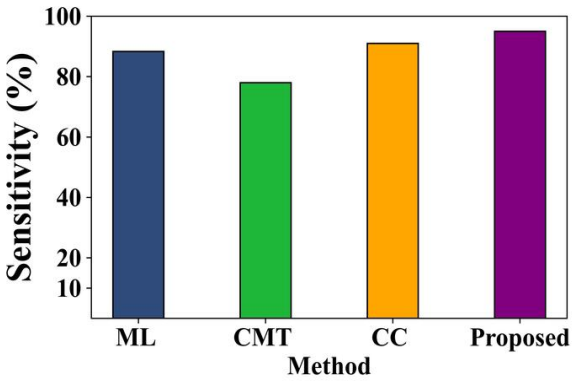

Figure 7. sensitivity.

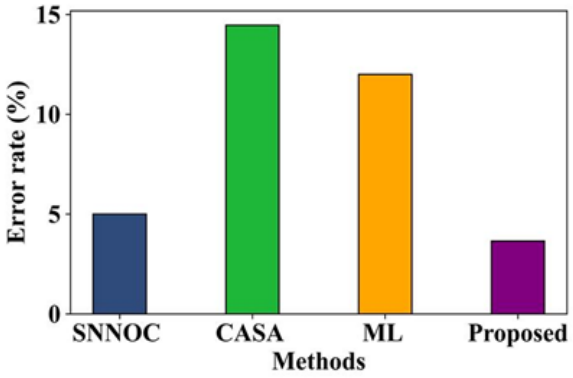

Figure 8. Error rate.

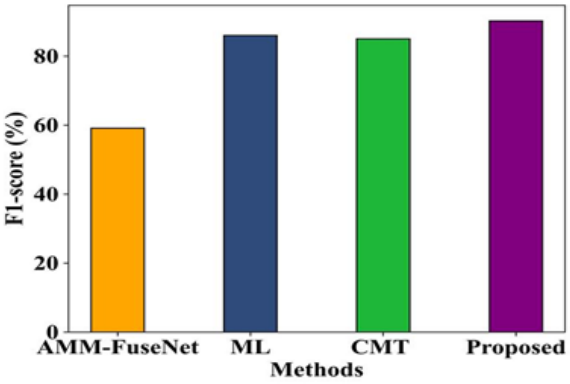

Figure 9. F1-score.

Figure 8delves into error rates, serving as a robust performance metric. The suggested approach consistently outperforms, SNNOC [18], CASA [20], and ML [23], with significant superiority observed, achieving notable error rate reductions of 26.01%, 54.01%, and 44.58%, respectively, compared to existing approaches.The F1-score, a unified metric incorporating precision and recall, evaluates a system's efficiency in correctly identifying true positives with reduced false positives. Comparisons with AMM-FuseNet [21], ML [23], and CMT [24] demonstrate the suggested approach's higher F1-score. At 100 epochs, significant improvements are observed, with increases of 36.66%, 4.66%, and 5.83% compared to existing methods, as depicted in Figure 9.

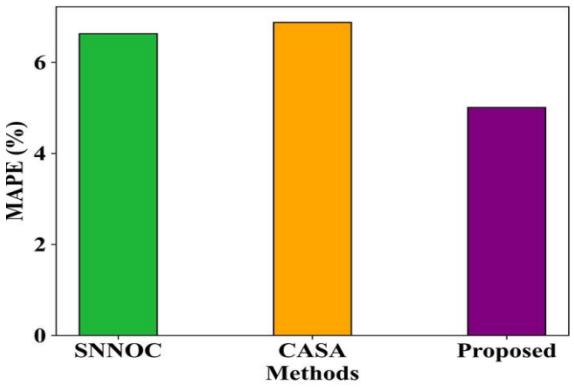

Figure 10. MAPE.

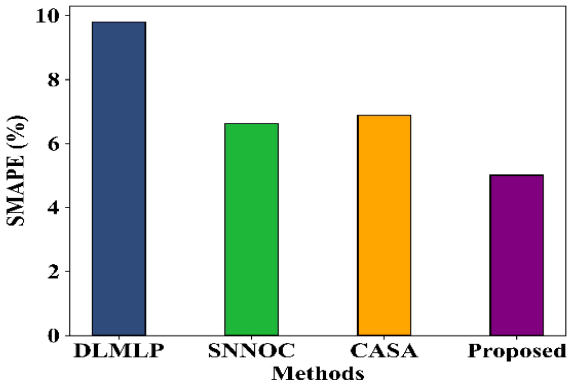

Figure 11. SMAPE.

Figure 10 visualizes the MAPE analysis for CYE in RSI, revealing the suggested approach's superior performance with a 24.43%, and 27.18% reduction in MAPE compared to SNNOC [18], and CASA [20], respectively. This metric is crucial for accurate prediction evaluation, especially effective in handling data with seasonal patterns.Figure 11 depicts the SMAPE analysis, emphasizing the suggested approach's reduced SMAPE with substantial improvements of 19.29%, 36.11%, and 27.18% at 100 epochs when compared to DLMLP [16], SNNOC [18], CASA [20], respectively, in CYE using RSI.

4.5. NDVI: A Measure of Wheat Plant Health Over Time

The Figure 12 depicts a line graph of the NDVI for wheat plants, tracking plant health during the growing season. It shows rising NDVI until flowering, signalling active growth, followed by a decline as plants mature. Key growth stages are marked, aiding in monitoring crop development.

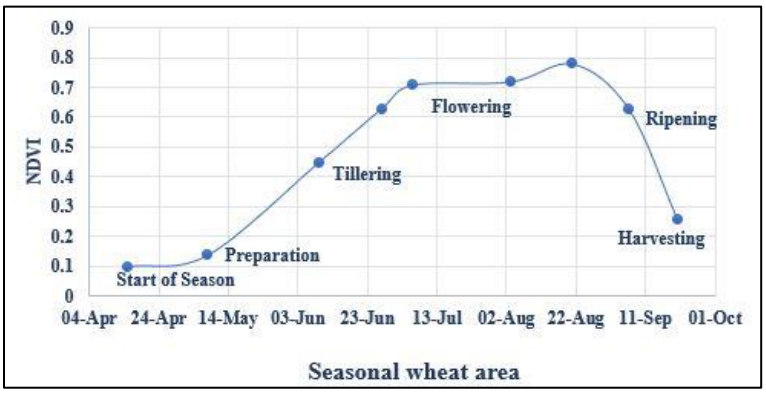

Figure 12. NDVI graph

The NDVI profile, influenced by weather, nutrients, and pests, serves as a crucial tool for crop monitoring and issue identification. It depicts phases like tillering, flowering, ripening, and harvesting. Following the tillering process, the plant undergoes growth, and there is a subsequent period before it begins to flower. This pattern aids in monitoring and estimating rice crop yield.

4.6. Crop Yield Prediction Performance Comparison

The table serves as a valuable reference for assessing and comparing the effectiveness of diverse methods in predicting crop yields based on the presented statistical metrics.

Table 5. Crop yield prediction performance comparison

METHODS	CROP	PARAMETERS	
		R^2	RMSE
DL framework under attention mechanism [1]	Wheat	0.63	5.02
Multi-head self-attention network [6]	Soybean, Wheat, Corn	-	7.8654
3-D Convolutional Neural Networks and Multi kernel Gaussian Process [10]	Wheat	0.80	0.73
Multimodal data fusion and DL [14]	Soybean	0.9	3.47
DL multi-layer perceptron [16]	Wheat	0.683	1.24
Simple algorithm for yield estimation [19]	Wheat	0.7	0.83
Carnegie-Ames-Stanford Approach and ML [20]	Wheat	0.554	3.36
Deep Fusion with Attention Neural Network [Proposed]	Wheat	0.95	0.65

Table 5 illustrates a performance comparison in predicting crop yields.The suggested method proves to be the most accurate method for predicting crop yield among those listed in the table. This is evident from its prominent R^2 value of 0.95 and the lowest RMSE value of 0.65. These metrics emphasize the close alignment between predicted and actual crop yields, surpassing the performance of alternative methods.

4.7. Discussion

MHCA-CEVN model for crop yield estimation is proposed, which significantly advances agricultural analytics by integrating sentinel-1 and sentinel-2 satellite data with field measurements. Proposed model, utilizing advanced preprocessing techniques like enhanced guided multi-layer side window box filtering and non-subsampled shearlet transforms, achieves an impressive accuracy of 97.59%, sensitivity of 95.21%, F1 -score of 90.21%, MAPF of 5.01% and an error rate of just 6.65%. These metrics surpass existing methods, demonstrating superior performance. The model's deep visual attention-based fusion and hybrid feature selection ensure a more comprehensive and precise estimation, offering substantial benefits for resource optimization and sustainable farming practices. Future work will expand its application and refine additional parameters to enhance its predictive capabilities.

The MHCA-CEVN model in proposed research shows remarkable efficacy in crop yield estimation by integrating remote sensing data with soil health parameters. Although proposed research work focuses on a specific crop and soil type, the model's methodology is highly adaptable for other crops and soils. By retraining the model with remote sensing and soil health data specific to new crops and soil types, and employing advanced preprocessing and hybrid feature selection techniques, the model will capture the unique characteristics of the new agricultural context in future experiment. Calibration with ground truth data ensures predictive accuracy, while iterative improvement based on validation results allows for fine-tuning. The model's scalability and transferability are evaluated across different geographic and climatic conditions, ensuring broad applicability. Customization for specific crop and soil challenges, such as varying growth cycles and nutrient needs, further enhances its versatility, making the MHCA-CEVN model a robust tool for diverse agricultural management scenarios and CYE.

5. CONCLUSIONS

This research work explores deep fusion with attention neural networks for soil health-based CYE through RSI. The results showcase the methodology's efficacy in capturing intricate relationships between soil health parameters and crop yield. Attention mechanisms, especially in the fusion process, enhance prediction accuracy by focusing on pertinent features. The deep fusion approach seamlessly integrates multi-modal information from remote sensing data, offering a holistic understanding of the agricultural landscape. The attention neural network adeptly weighs information from diverse sources, emphasizing the significance of soil health factors in CYE. This not only refines predictions but aligns with sustainable agriculture goals by highlighting soil health importance. RSI provide a non-intrusive, scalable solution for large-scale agricultural monitoring, promising precision agriculture insights. While a substantial advancement, future research avenues include exploring additional soil health parameters and advanced attention mechanisms for diverse crops and environments, along with investigating scalability and transferability for broader geographical applications. In essence, the presented deep fusion with attention neural network approach holds great potential for advancing CYE methodologies, contributing to technology's role in sustainable and efficient agricultural practices. Future work will focus on refining predictive capabilities by incorporating additional soil health parameters and advanced preprocessing techniques. Retraining the model with data specific to new crops and soil types, exploring advanced attention mechanisms, and assessing scalability for broader contexts will enhance CYE methodologies and support sustainable agricultural practices.

DATA AVAILABILITY STATEMENT

Data will be made available on reasonable request.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] Tian, H., Wang, P., Tansey, K., Han, D., Zhang, J., Zhang, S., & Li, H. (2021). A deep learning framework under attention mechanism for wheat yield estimation using remotely sensed indices in the Guanzhong Plain, PR China. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102375. doi:10.1016/j.jag.2021.102375

[2] Feng, Q., Yang, J., Liu, Y., Ou, C., Zhu, D., Niu, B.,& Li, B. (2020). Multi-temporal unmanned aerial vehicle remote sensing for vegetable mapping using an attention-based recurrent convolutional neural network. *Remote Sensing*, *12*(10), 1668. doi:10.3390/rs12101668

[3] Wang, H., Zhang, L., & Zhao, J. (2023). Application of a fusion attention mechanism-based model combining bidirectional gated recurrent units and recurrent neural networks in soil nutrient content estimation. *Agronomy*, *13*(11), 2724. doi:10.3390/agronomy13112724

[4] Ofori-Ampofo, S., Pelletier, C., & Lang, S. (2021). Crop type mapping from optical and radar time series using attention-based deep learning. *Remote Sensing*, *13*(22), 4668. doi:10.3390/rs13224668

[5] Li, Z., Chen, G., & Zhang, T. (2019). Temporal attention networks for multitemporal multisensor crop classification. *IEEE Access*, *7*, 134677–134690. doi:10.1109/access.2019.2939152

[6] Guo, Z., Chamberlin, J., & You, L. (2023). Smallholder Maize yield estimation using satellite data and machine learning in Ethiopia. *Crop and Environment*, *2*(4), 165–174. doi:10.1016/j.crope.2023.07.002

[7] Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, *14*(9), 1990. doi:10.3390/rs14091990

[8] Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sensing of Environment*, *259*, 112408. doi:10.1016/j.rse.2021.112408

[9] Hara, P., Piekutowska, M., & Niedbała, G. (2021). Selection of independent variables for crop yield prediction using artificial neural network models with Remote Sensing Data. *Land*, *10*(6), 609. doi:10.3390/land10060609

[10] Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Tian, Z., & Guo, H. (2021). Exploiting hierarchical features for crop yield prediction based on 3-D convolutional neural networks and multikernel gaussian process. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 4476–4489. doi:10.1109/jstars.2021.3073149

[11] Wang, D., Cao, W., Zhang, F., Li, Z., Xu, S., & Wu, X. (2022). A review of deep learning in Multiscale Agricultural Sensing. *Remote Sensing*, *14*(3), 559. doi:10.3390/rs14030559

[12] Kamarudin, M. H., Ismail, Z. H., & Saidi, N. B. (2021). Deep Learning Sensor Fusion in Plant Water Stress Assessment: A Comprehensive Review. *Applied Sciences*, *11*(4), 1403. doi:10.3390/app11041403

[13] Cai, W., Wei, Z., Liu, R., Zhuang, Y., Wang, Y., & Ning, X. (2021). Remote Sensing image recognition based on multi-attention residual fusion networks. *ASP Transactions on Pattern Recognition and Intelligent Systems*, *1*(1), 1–8. doi:10.52810/tpris.2021.100005

[14] Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and Deep Learning. *Remote Sensing of Environment*, *237*, 111599. doi:10.1016/j.rse.2019.111599

[15] Gao, L., Liu, H., Yang, M., Chen, L., Wan, Y., Xiao, Z., & Qian, Y. (2021). Stransfuse: Fusing swin transformer and convolutional neural network for Remote Sensing Image Semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 10990–11003. doi:10.1109/jstars.2021.3119654

[16] Tripathi, A., Tiwari, R. K., & Tiwari, S. P. (2022). A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation. *International Journal of Applied Earth Observation and Geoinformation*, *113*, 102959. doi:10.1016/j.jag.2022.102959

[17] Seydi, S. T., Amani, M., & Ghorbanian, A. (2022). A dual attention convolutional neural network for crop classification using time-series sentinel-2 imagery. *Remote Sensing*, *14*(3), 498. doi:10.3390/rs14030498

[18] Ali, A. M., Savin, I., Poddubskiy, A., Abouelghar, M., Saleh, N., Abutaleb, K.,& Dokukin, P. (2021). Integrated method for rice cultivation monitoring using sentinel-2 data and Leaf Area index. *The Egyptian Journal of Remote Sensing and Space Science*, *24*(3), 431–441. doi:10.1016/j.ejrs.2020.06.007

[19] Ma, C., Liu, M., Ding, F., Li, C., Cui, Y., Chen, W., & Wang, Y. (2022). Wheat Growth Monitoring and yield estimation based on remote sensing data assimilation into the safy crop growth model. *Scientific Reports*, *12*(1). doi:10.1038/s41598-022-09535-9

[20] Meraj, G., Kanga, S., Ambadkar, A., Kumar, P., Singh, S. K., Farooq, M.,& Sahu, N. (2022). Assessing the yield of wheat using satellite remote sensing-based machine learning algorithms and simulation modeling. *Remote Sensing*, *14*(13), 3005. doi:10.3390/rs14133005

[21] Ma, W., Karakuş, O., & Rosin, P. L. (2022). Amm-FuseNet: Attention-based Multi-Modal Image Fusion Network for land cover mapping. *Remote Sensing*, *14*(18), 4458. doi:10.3390/rs14184458

[22] Uribeetxebarria, A., Castellón, A., & Aizpurua, A. (2023). Optimizing wheat yield prediction integrating data from sentinel-1 and sentinel-2 with CatBoost algorithm. *Remote Sensing*, *15*(6), 1640. doi:10.3390/rs15061640

[23] Yeasin, M., Haldar, D., Kumar, S., Paul, R. K., & Ghosh, S. (2022). Machine learning techniques for phenology assessment of sugarcane using conjunctive SAR and optical data. *Remote Sensing*, *14*(14), 3249. doi:10.3390/rs14143249

[24] Trivedi, M. B., Marshall, M., Estes, L., de Bie, C. A. J. M., Chang, L., & Nelson, A. (2023). Cropland mapping in tropical smallholder systems with seasonally stratified sentinel-1 and sentinel-2 spectral and textural features. *Remote Sensing*, *15*(12), 3014. doi:10.3390/rs15123014

[25] Singha, C., Swain, K. C., & Jayasuriya, H. (2022). Growth and yield monitoring of potato crop using sentinel-1 data through cloud computing. *Arabian Journal of Geosciences*, *15*(19). doi:10.1007/s12517-022-10844-6

[26] Kong, W., Miao, Q., Liu, R., Lei, Y., Cui, J., & Xie, Q. (2022). Multimodal medical image fusion using gradient domain guided filter random walk and side window filtering in Framelet Domain. *Information Sciences*, *585*, 418–440. doi:10.1016/j.ins.2021.11.033

[27] Moad, M. S., Kafi, M. R., & Khaldi, A. (2022). A non-subsampled shearlet transform based approach for Heartbeat sound watermarking. *Biomedical Signal Processing and Control*, *71*, 103114. doi:10.1016/j.bspc.2021.103114

[28] Abdel-Basset, M., Mohamed, R., Zidan, M., Jameel, M., & Abouhawwash, M. (2023). Mantis search algorithm: A novel bio-inspired Algorithm for Global Optimization and Engineering Design problems. *Computer Methods in Applied Mechanics and Engineering*, *415*, 116200. doi:10.1016/j.cma.2023.116200

[29] Sarjamei, S., Massoudi, M. S., & Sarafraz, M. E. (2022b). Frequency-constrained optimization of a real-scale symmetric structural using Gold Rush algorithm. *Symmetry*, *14*(4), 725. doi:10.3390/sym14040725

[30] Sarjamei, S., Massoudi, M. S., & Sarafraz, M. E. (2022a). Frequency-constrained optimization of a real-scale symmetric structural using Gold Rush algorithm. *Symmetry*, *14*(4), 725. doi:10.3390/sym14040725

[31] Li, S., Wang, H., Zhang, C., & Liu, J. (2022). A self-attention feature fusion model for Rice Pest Detection. *IEEE Access*, *10*, 84063–84077. doi:10.1109/access.2022.3194925

[32] Kaur, A., Goyal, P., Rajhans, R., Agarwal, L., & Goyal, N. (2023). Fusion of multivariate time series meteorological and static soil data for multistage crop yield prediction using multi-head self attention network. *Expert Systems with Applications*, *226*, 120098. doi:10.1016/j.eswa.2023.120098

[33] Karthik, R., Joshua Alfred, J., & Joel Kennedy, J. (2023). Inception-based global context attention network for the classification of Coffee Leaf Diseases. *Ecological Informatics*, *77*, 102213. doi:10.1016/j.ecoinf.2023.102213

[34] Kong, J.-L., Fan, X.-M., Jin, X.-B., Su, T.-L., Bai, Y.-T., Ma, H.-J., & Zuo, M. (2023). BMAE-net: A data-driven weather prediction network for Smart Agriculture. *Agronomy*, *13*(3), 625. doi:10.3390/agronomy13030625

[35] Munaganuri, R. K., & Rao, Y. N. (2024). PAMICRM: Improving Precision Agriculture through multimodal image analysis for crop water requirement estimation using multidomain remote sensing data samples. *IEEE Access*, *12*, 52815–52836. doi:10.1109/access.2024.3386552

[36] Shao, G., Han, W., Zhang, H., Zhang, L., Wang, Y., & Zhang, Y. (2023). Prediction of maize crop coefficient from UAV multisensor remote sensing using machine learning methods. *Agricultural Water Management*, *276*, 108064. doi:10.1016/j.agwat.2022.108064

[37] Yang, S., Li, L., Fei, S., Yang, M., Tao, Z., Meng, Y., & Xiao, Y. (2024). Wheat yield prediction using machine learning method based on UAV Remote Sensing Data. *Drones*, *8*(7), 284. doi:10.3390/drones8070284

[38] Ramzan, Z., Asif, H. M., Yousuf, I., & Shahbaz, M. (2023). A multimodal data fusion and deep neural networks-based technique for tea yield estimation in Pakistan using satellite imagery. *IEEE Access*, *11*, 42578–42594. doi:10.1109/access.2023.3271410

AUTHORS

Vijaykumar P. Yele is a Research Scholar in the Department of Electronics & Telecommunication Engineering at Thakur College of Engineering and Technology, Mumbai. His areas of specialization include networking & data compression, image processing, machine learning and AI.

R. R. Sedamkar is a Professor in the Computer Engineering at Thakur College of Engineering and Technology, Mumbai. His areas of specialization are networking & data compression, image processing, visual cryptography, machine learning and AI.

Sujata Alegavi is an Associate Professor in the Department of Internet of Things at Thakur College of Engineering and Technology, Mumbai. His specialization areas are communication networks, image processing & video processing and remote sensing.