Enhancing Crop Yield Prediction Through Image-Driven Multi-Level Feature Learning and Regularized SNN-GRU

Senthil Kumaran Vijayalakshmi Natarajan[1], Rajkumar Ganesan[2], Karthikeyan Sakthivel[1],
Gayathri Devi Thirugnanasambandam[3*]

[1] Department of ECE, SRM Institute of Science and Technology, Tiruchirappalli 621105, India
[2] School of Electrical & Electronics Engineering, SASTRA Deemed to be University, Thanjavur 613401, India
[3] Department of ECE, Srinivasan Ramanujan Centre, SASTRA Deemed University, Kumbakonam 612001, India

Corresponding Author Email: devigayathri77@src.sastra.edu

ABSTRACT

A Regularized Spiking Neural Network and Gated Recurrent Unit (RSNN-GRU) with multi-level feature learning are used in the proposed model to forecast agricultural yield. Using t-SNE and Kernel PCA approaches, the preprocessing step encompasses data cleaning, standardization, normalization, and dimensionality reduction. The dataset is divided into training and testing sets with an 80-20 split to create the model. Statistical and higher-order statistical features as well as two newly proposed features, adaptive weighted kurtosis, and adaptive weighted skewness, are all used in feature extraction. A Self-Adaptive Farmland Fertility Optimization (SAFFO) algorithm is used for feature selection, improving feature selection performance. The SAFFO algorithm's hyperparameter adjustment helps the RSNN-GRU model even more. The performance of the proposed model is compared to that of previous deep-learning models for agricultural yield prediction using some evaluation metrics. The proposed model performs better than others, according to the results, and has a lot of potential for usage in the agricultural sector. Python 3.7.9 is used to carry out the proposed model's implementation. The results show that the proposed SNN-GRU has better performance in terms of predicting the better crop-yield with an accuracy of 96.1% and 96.5% with the help of hyperparameter learning rate of 0.7 and 0.8 respectively. The other metrics such as precision, sensitivity, specificity and F-measure have shown better results when compared to existing models such as CNN, RNN, LSTM and DNN.

1. INTRODUCTION

Crop yield prediction is the practice of estimating how many crops will be harvested from a specific area or farm utilizing a variety of technology and data analysis techniques. Farmers and agricultural specialists use this information to make judgments about crop management, resource allocation, and market forecasts. Predicting crop yields accurately can aid in streamlining farming procedures, boosting productivity, and ultimately enhancing food security. based on remote sensing, agricultural yield. Due to the lack of yield mapping devices among farmers, machine learning methodologies for prediction are limited [1], and enhancing the accuracy of agricultural yield prediction can be done by integrating MLR coefficients and bias in ANN [2]. For all built ML models, adding fresh inputs from a cropping systems simulation model (APSIM) can increase the precision of yield forecasts [3]. It is critical for the economy, policymaking, and increasing global food security measures to anticipate soybean production in Brazil [4]. Owing to its ubiquity, accurate crop production forecasts and yield loss estimation owing to soil salinity are essential in Mediterranean countries [5]. At the county level, the CNN-LSTM model is used to forecast soybean production both at the end of the season and during the growing season. Three Self-Organizing Map models have also been utilized to link precision agriculture data with yield productivity is frequency classes, including CPANN, SKN, and XYF [6, 7]. A novel method for predicting agricultural yield by capturing hierarchical information is proposed utilizing a 3-D convolutional neural multi-kernel network, overcoming the complexity of crop growth and enhancing performance [8]. The efficiency of machine learning approaches for silage maize yield prediction is limited, especially when cultivated at various times and fields within a region, although they have been employed for crop monitoring and yield prediction using remotely sensed data. During the procedure, an electronic instrument is used to record the relative humidity, ambient temperature, and geographic coordinates, including latitude and longitude [9, 10].

The development of crop productivity prediction techniques can help farmers and other stakeholders choose the right crops and agronomic procedures for various climatic situations [11]. The goal of the study was to evaluate the effectiveness of combining data from a DNN with multispectral, UAV-mounted RGB, and thermal meters to predict soybean grain production [12]. Data analysis and computational tools must

be developed to satisfy the needs of farmers to help them make decisions and get insights [13]. Crop yield can be optimized by using machine learning algorithms to extract meaningful information from precision agriculture equipment [14]. On the Google Earth Engine (GEE) platform, a modeling framework has been developed to forecast winter wheat yield. It integrates data on the weather, remote sensing, and soil [15]. Crop yield prediction plays a significant role in better allocation of resources, ensuring food security, and data-driven decision-making in the farm. Conventional machine learning and deep learning models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) models, have been extensively used for yield prediction. However, these approaches have several drawbacks. Feature extraction in conventional techniques struggle to capture complex statistical relationships in agricultural data, as the process relies mainly on raw data. Hence the feature extracted are sub-optimal which may lead to incorrect predictions. Principle Component Analysis (PCA) fails to preserve the relationship between input variables which impacts predictive performances. Inefficient feature selection leads to increased computational costs. To address these limitations, we propose a Regularized Spiking Neural Network with Gated Recurrent Unit (RSNN-GRU) model with multi-level feature learning which integrates.

1. Adaptive weighted kurtosis & adaptive weighted skewness is used to capture the higher order statistical properties of crop yield data for better accuracy then conventional learning models

2. Self-Adaptive Farmland Fertility Optimization (SAFFO) Algorithm improves feature selection and hyperparameter tuning to enhance the model's efficiency

3. To reduce computation complexity, Two-Fold Dimensionality Reduction (t-SNE and KPCA) ensures retaining discriminative features and information.

There are various components to this study article. A literature review is provided in Section II. In Section III, the putative mechanism employed in this work is outlined, and in Section IV, the experimental findings are covered. Section V provides an overview of the research's conclusions.

2. LITERATURE REVIEW

To forecast rice grain yield in 2017, Zhou et al. [16] employed single-stage and multi-temporal vegetation indices generated from multispectral and digital photos. According to the findings, regardless of the type of imaging, the booting stage is the best stage for prediction utilizing VIs at a single stage. A cutting-edge UAV-based hyperspectral push broom scanner system that has been created to gather superior spectral and spatial quality high-resolution data was proposed in 2018 by Kanning et al. [17]. This revolutionary technique was best suited for exploring smaller areas, such as specific plots in precision farming. Zhao et al. [18] explored the potential application of S2 MSI imagery for mapping the 2020 dryland wheat yield at the field scale and across the NEAUS. S2 time series were used to derive vegetation indices (VIs) and vegetation growth metrics. Then, the accuracy of these traits in predicting wheat yields was tested. At the shire scale, we evaluated whether merging VIs with a crop stress index produced by the Oz-wheat model improved yield prediction. Wei et al. [19] suggested partitioning soybean evapotranspiration for yield projections. The dual crop coefficient method accounting for transpiration and soil evaporation components was used in the SIMDualKc water balance model. This allowed for the full analysis of soybean water use, an important step towards understanding yield gaps and enhancing crop management practices. Cai et al. [20] have presented a range of machine learning and regression algorithms for predicting wheat yields in Australia based on weather and satellite data in 2019 using Exploratory Data Analysis (EDA).

According to the study, accurate yield forecasts are better obtained when one uses satellite and climate data information. Wan et al. [21] proposed to combine structural and spectral information produced from UAV-based images in order to improve the prediction of rice yield across growing cycle in 2020. The aim of the study was to increase the accuracy and consistency of meteorological data used in yield forecasting models by using both types of data. In 2018, Sabzi et al. [22] proposed a computer vision-based expert system for site-specific spraying to identify potato plants and three types of weeds (*Secale cereale L., Polygonum aviculare L., and Xanthium strumarium L.*) This technique employs advanced image processing algorithms to reliably differentiate between desirable plant species and weeds, enabling the same crop health to be maintained with less herbicide. Saranya and Nagarajan [23] predicted that low-resolution satellite images would be required in operational systems in 2020, because they were being widely used for crop monitoring and yield prediction. Due to their wide coverage and high temporal frequency, these images are a profitable alternative at both national and regional scales. This capability makes them widely used for numerous agricultural applications such as crop health monitoring and forecasting yield potential. Elavarasan and Vincent [24] proposed a new hybrid regression-based generalization method called Reinforcement Random Forest in 2021, which outperformed classical machine learning models, which conventional machine learning algorithms like decision trees, random forests, and artificial neural networks and gradient boosting and DQNs. This novel strategy dramatically boosts performance through reinforcement learning to maximize utilization of available samples during the tree construction stage. In 2021, Gong et al. [25] proposed a greenhouse crop yield prediction method based on fusing two advanced frameworks, namely temporal convolutional network (TCN) and recurrent neural network (RNN), to handle the temporal sequence data from the greenhouse environment. This approach produces more accurate and reliable yield estimates.

2.1 Problem statement

The process of collecting the evaluation data from UAV-based multispectral and digital images, as well as the processing of images in multi-temporal vegetation indices, takes a long time and is resource-intensive, thus the application of multi-temporal vegetation indices for prediction of rice production using UAV-based multispectral and digital images is limited. External factors, like weather, affect crop growth and development and also influence the accuracy of projections [17]. A dataset of unmanned aerial vehicle multispectral images acquired over a field [26] is considered for this research. A limitation of this approach is that it requires the use of specialised equipment and knowledge to fly the UAV and analyse the data, particularly: (1) multi-temporal UAV-based RGB and multispectral images for predicting rice

grain yield; (2) UAV images of multi-temporal bases; (3) UAV multispectral images. In addition, how different environments, growth stages and climate in different locations would affect the population models limited the applicability of it [22]. Another disadvantage is that the implementation of multi-temporal UAV-based RGB and multispectral images for prediction of rice grain yield requires advanced equipment and qualified personnel that can operate the UAV and interpret the data. In addition, differences in environmental conditions and growth stages between geographical sites and seasons may restrict the applicability of the model [23]. On the downside, agricultural output prediction systems may be based on data from the past, which does not account for rapid changes or unanticipated events, such as diseases or extreme weather. In addition, variations in environmental conditions, soil quality, etc., which could potentially impact crop growth and development, may also be negative [24]. Crop yield prediction algorithms may behind aggressive and exact data inputs, such as weather and dirt data, that aren't always to be had or correct, which may be a downside. Additionally, prediction accuracy can be affected by changes in weather, outbreaks of diseases, pest infestations, or other factors that may influence the growth and development of a crop.

3. PROPOSED METHODOLOGY

The proposed method for agricultural production prediction consists of several procedures aimed at obtaining precise and efficient estimates. First, an open-source platform allowed one to compile a dataset on agricultural yields. Pretreatment processes like data cleansing, standardization, normalizing, two-fold dimensionality reduction employing t-SNE and KPCA help to provide a clean dataset. Features from the data are extracted using statistical and higher-order statistical features as well as the proposed adaptive weighted kurtosis and skewness features following the division of the data into training and testing sets. Farmland Fertility Optimization (FFO) is a recently developed self-adaptive metaheuristic algorithm used for feature selection. The FFO algorithm is changed to improve performance and becomes self-adaptive with changing its parameters. Better optimization solutions and feature selection performance are obtained by the proposed Self- Adaptive FFO algorithm. The top features found by feature selection then feed the classifier forecasting crop yield.

Agricultural yield is estimated using a modified spiking neural network (MSNN), which features a regularized loss function to improve prediction performance. To enhance the MSNN, a gated recurrent unit (GRU) is included into its architecture and produces a Regularized SNN with GRU (RSNN-GRU). Hyperparameter adjustment is done using the SAFFO approach, therefore optimizing the RSNN-GRU model even more. The performance of the proposed model is finally assessed using multiple criteria including RMSE, MAE, BCE, CCE, F1-Measure, and Accuracy. To evaluate the suggested approach, several well-known models including "Convolution Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Deep Neural Network (DNN) are compared. It is a feasible option for estimating agricultural output since the proposed technique performs better and generates more accurate forecasts. Figure 1 shows the whole flow of the suggested model.

3.1 Pre-processing

Many pre-processing techniques can help to clean a dataset before classification modelling. These comprise standardizing the data to assure homogeneity, normalizing the data to take into account different scales, and data cleansing to eliminate disparities. T-SNE and kernel PCA are two methods for two-fold dimensionality reduction that can help to lower the required number of variables for analysis.

3.1.1 Data cleaning
During data preparation, data is examined, cleaned, and corrected to create a wholesome, error-free final data set. In a wide range of applications, faulty or incomplete data can lead to poor analysis and incorrect conclusions with dire implications. In order to prepare your data for analysis you will need to perform a data cleaning task, which means detecting and removing any inaccuracies or inconsistencies (e.g., duplicate data, missing values, problems with format). However, cleaning the data you can make sure the dataset is accurate, reliable, and suitable for analytics, and that any derived conclusions are legitimate and useful.

3.1.2 Data standardization
Data standardizing means transforming data into a format that is quickly understood and interpret by computers. This enables data analysis to be performed in a simplified manner, making it possible for a range of systems to successfully exchange data. Data standardizing help to keep the data correct and quality. Uniform data provides a backing for the decision-maker to recognize glitches and aberrances. Also, data standardizing saves time and resources while working on data administration and help to achieve better communication between various teams and companies that utilize the same data.

3.1.3 Data normalization
Data normalization helps to eliminate redundancy and dependency issues therefore guaranteeing consistency and accuracy in data. It means breaking down a large database into smaller, more controllable tables with individual sets of relevant data. This helps to reduce the likelihood of data discrepancies that is, duplicate or inconsistent data. Normalization also helps to minimize data duplication, so producing an easier to update and maintain database with less storage space usage.

3.1.4 Two-fold dimensionality reduction
Dimensionality can be lowered using two different approaches: Kernel Principal Component Analysis (KPCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE). These techniques can help to reduce the feature count of a dataset, therefore enabling analysis and visualization.

3.1.5 t-Distributed Stochastic Neighbor Embedding (t-SNE)
Using the efficient data visualization method called t-Distributed Stochastic Neighbor Embedding (t-SNE), high-dimensional data can be effectively reduced to two or three dimensions for visualization. When presenting complex and erratic data structures, the method is particularly useful. Using t-SNE, data points are first evaluated in a high-dimensional space for pairwise similarities and subsequently a low-dimensional representation of the data is optimised maintaining these similarities. t-SNE is useful for high-

dimensional data representation visualization as it preserves local structures. t-SNE retains nonlinear structures inherent in remote sensing and soil data unlike PCA which assumes linearity. It is useful for identifying clusters in crop yield data that may correspond to particular weather conditions, soil types, or farming practices.

3.1.6 KPCA

A dimensionality reduction approach called KPCA converts high-dimensional data into a lower-dimensional space. The KPCA is a nonlinear generalization of PCA through kernel functions. The advantage of utilizing a kernel function to translate the data to a high-dimensional space before performing PCA makes it similar to the conventional PCA technique. When the data cannot be separated linearly, or when the connection between the features is nonlinear, KPCA may be helpful. KPCA can find complicated associations between the characteristics that would be impossible to find with conventional PCA by mapping the data into a high-dimensional space. This is especially beneficial when trying to analyze heterogeneous agricultural datasets, wherein attributes such as soil composition, weather parameters, NDVI indices, etc. have complex interdependencies. Feature extraction, which entails changing the data into a new collection of features that are more beneficial for downstream analysis, is another application of KPCA. KPCA retains important spatial and temporal variations in the ability to predict the yield of agricultural production, unlike ICA that aims of statistical independence.

3.2 Feature extraction

Higher-order statistical features like skewness, variance, and kurtosis are frequently utilized in feature extraction together with more traditional statistical features such as Harmonic Mean (HM), median, and standard deviation. To extract more significant characteristics from data, the proposed techniques of adaptive weighted kurtosis and adaptive weighted skewness can be used.

3.2.1 Statistical features

Statistical features are foundational in data science and analytics because they provide insight into the distribution data and inform the selection of the best machine learning models. These characteristics include statistical properties such as the mean, median, and percentiles. Statistically based feature selection methods are also of utmost importance in the analysis of the data since they help discern the most important features that are impacting the target variable significantly. Methods make models more accurate and efficient by removing unnecessary or irrelevant traits. The study of statistical correlations between input and target variables is a fundamental aspect of machine learning model optimization.

- HM: HM is a statistical measurement which is the reciprocal of the average of the reciprocal of the data values. It also accounts for all the observations and gives less weighting to the larger values as compared to the smaller ones. Hence, it is employed whenever there is a need to specify the nuances. When ratios are involved such as the case of average speed or price per unit the HM is used. This feature of balancing out the small and large numbers that can skew the entire view is how HM can help to give a better picture of actual data. The arithmetic mean is obtained by listing the objects reciprocally and dividing the individual counts from the overall count.

$$h_m = \frac{m}{\left[\left(\frac{1}{a_1}\right)+\left(\frac{1}{a_2}\right)+\dots+\left(\frac{1}{a_m}\right)\right]} \tag{1}$$

- Median: A median is a statistical estimate of central tendency that shows the midway value of a set of statistics. It is the point at which a set of data is split in half with half of the values above the median and the other half below. Usually called the 50th percentile, the median is the central observation in distribution. Calculated with m as the total number of observations, the mean and the median have value.
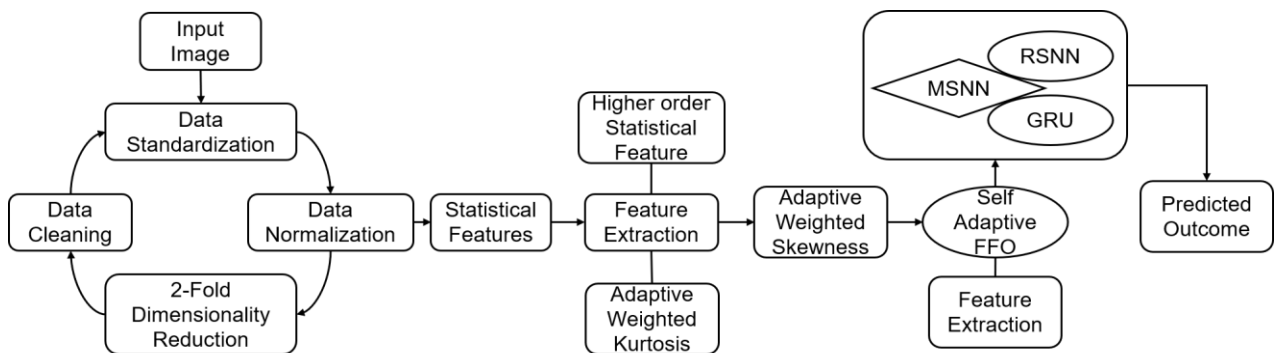
$$median = \frac{(m+1)}{2} \tag{2}$$

Figure 1. Proposed model for crop yield prediction

- Standard Deviation: The standard deviation is a statistical instrument for measuring dispersion or variance in a dataset. The square root of the variance shows the extent of data point deviations from the mean. Whereas a smaller standard deviation indicates a steadier distribution, a larger standard deviation indicates a more diverse sample. Since it clarifies the features of the dataset and helps to identify any outliers or patterns, this statistic is helpful in many fields, including risk analysis and quality management. The mean's, medians, and

standard deviation's respective values are found by:

$$S_d = \sqrt{\frac{1}{m-1}\sum_{p=1}^{m}(a_p - \bar{a})^2} \tag{3}$$

3.2.2 Higher order Statistical Features

- Skewness: Skewness is a statistical measure of a dataset's asymmetrical distribution. It implies that if the distribution spans one side of the peak longer than the other. A positive

skewness denotes a right-skewed distribution; a negative skewness denotes a left-skewed distribution.

$$Skewness = 3 \times \frac{(Mean-median)}{SD} \qquad (4)$$

• Variance: Variance in statistics is a measure of dataset distribution. It measures dispersion and assesses the fluctuation of data points about the mean. Data can be expressed both as a group or non-groups. Individual data points make up ungrouped data; grouped data is data shown as class intervals.

$$Var(A) = E[(A - \mu)^2] \qquad (5)$$

• Kurtosis: Kurtosis measures how much outliers are present in your data, the higher the kurtosis means that there is a high presence of outliers. It reflects how much the tails of a distribution deviate from a normal distribution. High excess kurtosis indicates the presence of more aggressive outliers, and it measures the tail of the distribution compared to a Gaussian distribution. Therefore, E = E(a).

$$\beta_2 = \left(\frac{E(a)^4}{(E(a)^2)^2}\right) - 3 \qquad (6)$$

3.2.3 Adaptive weighted kurtosis

Kurtosis which is a measure of tail data is helpful in determining outliers which is helpful in data related to crop yield for rare events. An example is adaptive weighted kurtosis, which gives relative weight to check the values more if there are any changes into data set. The adaptive weighted kurtosis Approach uses historical data to identify relationships between distributions and, ultimately, predicts crop yields by leveraging patterns it finds in the kurtosis of that distribution. This method allows for more specific estimations of crop yield by changing the weights given to different parts of the distribution as the shape of the distribution changes through time. The first step in predicting crop yield via adaptive weighted Kurtosis is gathering historical data about the crop yield in a specific location. It should be spread over multiple years and involve different climate and environment variables. Once the historical data is collected, the next step would be to calculate the kurtosis of the distribution of crop yields recorded in each year in the data set using the following formula:

$$Kurtosis = [\textstyle\sum(A_j - \bar{A})^4 / M] / (K^4) \qquad (7)$$

where, A is the mean of yield over all years, Aj is the yield in the j-th year, M is the number of years and K is the standard deviation of yield over the years. The weights for each year are then determined according to the kurtosis value after the kurtosis values have been obtained. This way, in the case of high kurtosis values, more weight can be placed on the years with a high possibility of extreme yields. If you want, you can calculate the weight using this formula:

$$Weight = (S - \bar{S}) / K \qquad (8)$$

where, S is the kurtosis in a given year, $\bar{S}$ is the average kurtosis and K is the average kurtosis standard deviation after the weights are determined, the adaptive weighted mean yield for the following year can be computed as a weighted mean of all the historical yield data. In the previous step, we established the weight for every single year. Where AWM is the adaptive weighted mean, μ is the mean and σ is the standard deviation.:

$$AWY = \textstyle\sum(Weight * B_j) / \sum Weight \qquad (9)$$

where, B_j is the particular year yield, Weight is the weight of the particular year as decided at the previous step, AWY is the adaptive weighted mean yield for future year. Adaptive weighted Kurtosis offers a way for farmers and policymakers to more accurately predict agricultural yields by taking into consideration variations in the crop yield distribution over time. This could lead to better decision-making and more accurate estimates of agricultural yields, which would drive towards improvements in crop management and higher yields. Using an adaptive version of the conventional kurtosis, this improved high impact events detection in crop yield data and achieved better prediction performance.

3.2.4 Adaptive weighted skewness

Skewness refers to the asymmetry in the distribution of data points and indicates whether yields are skewed upward or downward. The statistical method of adaptive weighted skewness can be applied to predict crop yield when using past agricultural production data and investigate the distribution of agricultural production data. Skewness is a measure of the asymmetry of a distribution can inform whether the distribution predicting crop yields is skewed low or high in value. In practice, the adaptive weighted skewness approach involves changing the weights assigned to each observation in the historical data set based on their distance from the weighted mean of the historical data set. Adaptive weighted skewness is formulated in terms of data weighted standard deviation of the observations and total number of observations in the data set. This will cover negative (or positive) yield since Adaptive Weighted Skewness is completely dependent on time- and location-based skewness and thus ensures an accurate assessment of yield. By slowly updating weights of observations based on changes in the distribution of yield data over time, adaptive weighted skewness can eventually help identify crop performance trends in a more realistic light. When combined with adaptive weighted kurtosis, which measures the peak of the distribution, adaptive weighted skewness can provide a more comprehensive view of the distribution of agricultural yields, and contribute considerations for forecasting crop yields.

$$AWS = \Sigma(w_j * (a_j - \bar{a})^3) / ((\Sigma(w_j))^3/2 * m * k^3) \qquad (10)$$

where,
w_j = weight of the j-th observation
a_j = j-th observation
$\bar{a}$ = weighted mean of the observations
m = number of observations
k = weighted standard deviation of the observations

Adaptive weighted skewness and adaptive weighted kurtosis can make farm and government decisions, including crop management practices, such as planting techniques, fertilizer application, insect control, etc. The use of machine learning methods also allows for previous distributions of yields and their characteristics over time to be factored into forecasts, ensuring greater accuracy in predictions that would allow for better decision making on the ground, which could ultimately lead to more sustainable agricultural practices.

3.3 Feature selection

A new self-adaptive metaheuristic algorithm called FFO is proposed to accomplish optimal feature selection. Image Segmentation Locks its Focusing Features into the FFO Algorithm Output: 1) Setting of the FFO algorithm has been changed to the self-adaptive types and improved its performance. We have presented Self-Adaptive FFO to better optimize the feature selection process using new alternatives. In the next step, the selected best features are taken as inputs for the classifier to predict the yield of a crop resulting in better agricultural practices.

3.3.1 Self-adaptive FFO

In order to obtain the optimal feature selection, a new self-adaptive metaheuristic algorithm has been introduced called FFO. The parameters of the FFO algorithm are improved to self-adaptive performance. Based on the above analysis, the Self-Adaptive FFO is proposed to enhance the procedure of feature selection with more effective optimization options. The selected best features are then used as input by the classifier to predict the crop yield which results in better agricultural practices.

Initializing FFO Parameters and Solutions

- Step 1: Upgradation of Soil Quality

In this step, the FFO algorithm generates the total population M_{pop} in the search space using the number of cropland segments (s) and the possible solutions in each piece (m). Initializing both FFO parameters and solutions serves not just as the first step of the algorithm but also opens doors for further inquiries.

$$M_{pop} = s.m \tag{11}$$

Pics of the FFO Algorithm-The algorithm doesn't use general equations like other algorithms, but generates the values of the solution from the supplied data range and its distribution. This equation, Eq. (12), is utilized for generating random solutions in the search space. This stage is crucial to ensuring a diverse solution and is instrumental in initializing the FFO parameters.

$$a_{ji} = LL_i + rand \cdot (UL_i - LL_i), \\ \forall j \in m, i \in s \tag{12}$$

where, LL_i and UL_i are respectively the min/max bounds of solution x and a rand is a number randomized between 1 and 0.

- Step 2: Define SQ in each section of farming

At this point, solutions in each piece of farmland can be determined, as indicated in the subsequent Eq. (13).

$$Portion_k = a(x_i), x = m \cdot (k-1) + 1 : m \cdot k, \forall k \in \\ s, i \in s \tag{13}$$

The average of the present solutions in each area of farmland is used to determine SQ.

$$Fit_{portion_k} = mean(allFit(a_{ji})inPortion_k), \\ \forall k \in s, j \in m, i \in s \tag{14}$$

- Step 3: Memory modification

The best answers for every section are saved in the global memory, along with a few of them in the accompanying regional memory for that section. The number of solutions in the regional memory N_{Reg} and the number of solutions in the universal memory N_{Univ}, respectively, are defined by the following Eqs. (15) and (16).

$$N_{Reg} = round(T \cdot m), 0.1 < T < 1 \tag{15}$$

$$N_{Univ} = round(T.M_{pop}), 0.1 < T < 1 \tag{16}$$

These memories retain solutions because they are relevant and suitable. At this point, both memories had been updated, and the greatest and worst parts had been determined. This stage involves defining the finest and worst parts when memories have been refreshed.

- Step 4: Change SQ in each section of cropland

The region with the lowest SQ will currently have the most variations, and its responses are coupled with a solution that is already stored in the global memory, as shown in the following Eq. (17):

$$A_{new} = q \cdot (A_{ji} - A_{N_{Univ}}) + A_{ji} \tag{17}$$

$$q = \alpha \cdot rand(-1.1) \tag{18}$$

where, $A_{N_{Univ}}$ is a parameter of FFO that is first adjusted between 0 and 1, and $A_{N_{Univ}}$ is randomly selected from the existing solutions within the universal memory, where A_{new} represents a new solution, A_{ji} is a solution in the poorest area of farmland that is chosen to perform variations, and $A_{N_{Univ}}$ is a parameter of FFO that is adjusted between 0 and 1. After the poorest area of farmland has undergone adjustments, the remaining areas must be integrated with the available options across the entire search width. The solutions that can be found in the remaining portions are defined by the following Eqs. (19) and (20):

$$A_{new} = q \cdot (A_{ji} - A_{vj}) + A_{ji} \tag{19}$$

$$q = \beta \cdot rand(0.1) \tag{20}$$

where, A_{vj} is an FFO parameter that is initially set between 0 and 1 and randomly selected from the already-found solutions across the whole search breadth.

- Step 5: Incorporate Soil

Every soil in the farming parts is now included under the best options in their local memory, which is called $Fitness_{Reg}$ the end. As shown in the following Eq. (21), the incorporation of all accessible solutions is not constrained by local memory or at this stage, but rather, they are integrated with the best solutions currently available to improve the quality of solutions in each segment.

$$Q = \begin{cases} A_{new} = A_{ji} + \omega_1 \cdot (A_{ji} - Fitness_{Univ}(y)), \forall H > rand \\ A_{new} = A_{ji} + rand \cdot (A_{ji} - Fitness_{Reg}(y)), \forall H \leq rand \end{cases} \tag{21}$$

where, H represents an FFO parameter that is initially adjusted and gradually decreased as a result of multiplying by a factor D_u based on algorithm reiteration as shown in the following Eq. (22) where 1 represents an FFO parameter that is initially adjusted and gradually decreased as a result of multiplying by a factor D_u based on algorithm reiteration.

$$\omega_1 = \omega_1 \cdot D_u, \qquad 0 < D_u < 1 \tag{22}$$

- Step 6: Finishing Conditions

When the FFA algorithm has examined all potential solutions in the search space and has reached the maximum number of optimization iterations, it stops.

MFFA

The FFA algorithm has been enhanced with MFFA to improve performance. The addition of a mutation factor and the inclusion of a local search mechanism, respectively, are the alterations performed in the fourth and fifth phases of traditional FFA. In the fourth stage. The Levy flight motion method was used in place of Eqs. (18) and (20). Levy flights, a random walk approach, are employed in this stage to alter Eqs. (18) and (20) for improved exploitation and exploration. To enhance the FFA method, the modified equations add the Levy flights technique, which involves taking a sequence of successive random steps.

$$q = \alpha \times Levy(M_{var}) \qquad (23)$$

$$q = \beta \times Levy(M_{var}) \qquad (24)$$

The MFFA including Levy flights enable the algorithm to have both local and global searching capabilities. In the exploitation and exploration phases of the algorithm, levy flights represent a series of successive random movements. Additionally, the same formula to determine Levy flights is also used in ways to optimize performance.

$$Levy(M_{var}) = 0.01 \times \frac{dd_1 \times \delta}{|dd_2|^{\frac{1}{\beta}}} \qquad (25)$$

The step size, d, in Eq. (25). In the random walk process, diffusion of step sizes is controlled by the Levy flight exponent. This means that high exponents increase chances of taking longer steps which is better for a global search, and low exponents increases chances of taking shorter steps which is better for a local search.

$$\delta = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}} \right) \qquad (26)$$

where, $\Gamma(a) = (a-1)!$.

In the fifth stage, the sine-cosine method adds the sine-cosine function to the Eq. (21), so that the various solutions can oscillate around the best one. The output equation appears in an altered version that includes sine-cos function and flooded with paths teaming to converge to the optimal.

$$Q = $$
$$\begin{cases} A_{new} = A_{ji} + d_1 \times \sin(d_2) \times \left(A_{ji} - Best_{Global}(y)\right).H > rand \\ A_{new} = A_{ji} + d_1 \times \cos(d_2) \times \left(A_{ji} - Best_{local}(y)\right).H > rand \end{cases} \qquad (27)$$

$$d_1 = 2 - It \times \left(\frac{2}{MaxIt}\right) \qquad (28)$$

$$d_2 = (2 \times pi) \times rand(0,1) \qquad (29)$$

where, "It" is the current iteration, MaxIt is the maximum number of iterations and "rand" is a random number generated uniformly over (0-1).

3.4 Crop yield prediction

The authors proposed a modified Spiking Neural Network (MSNN) with Regularized Loss and the Gated Recurrent Unit (GRU) to enhance agricultural production forecast. SAFFO while readjusting hyper-parameters (such as Batch Size, Momentum, Learning Rate, Epoch, etc.) equipped the resultant Regularized SNN with GRU (RSNN-GRU) model, and so far a significantly improved prediction performance was reached using criteria such as RMSE, MAE, BCE, CCE, F1-Measure, and Accuracy. RSNN-GRU achieving better accuracy underscores its potential as a tool for predicting crop yield. Figure 2 illustrates the layered architecture for the proposed RSNN and GRU.

3.4.1 Modified Spiking Neural Network

Regularized SNN: Crop yield prediction is important for food security and sustainable agriculture. The regularized Spiking Neural Network (RSNN), for instance, employs spiking neurons to account for the complex relationship between environmental factors and agricultural output. The RSNN employs a regularized loss function adding a penalty term to the context of the objective function to reduce overfitting and improve generalization performance. This gives our network and urge to have smaller weights. A regularization term like LT1 or LT2, for example, is added as a penalty that is inversely proportional to the sum of absolute values of the weights or the respective squared values. LT1 regularization also prefers sparse weights as it reduces network complexity, helps in overfitting, and increases generalization performance.

The RSNN can be trained using various approaches, such as backpropagation through time (BPTT) or spike-timing-dependent plasticity (STDP), by adjusting the network weights based on the difference between the desired and the actual outputs. This allows the network to predict more accurately the relationship between environmental condition and crop productivity. In conclusion, RSNN is a new method for predicting agricultural yield with optimal generalization performance since it has the ability of capturing nonlinear relationships, and the ability of overcoming the problem of overfitting in the model, and it is promising once again. A principal component of the RSNN, the regularization term of the loss function promotes simpler network architectures, enhancing prediction performance. It is stated that the regularized Loss function is
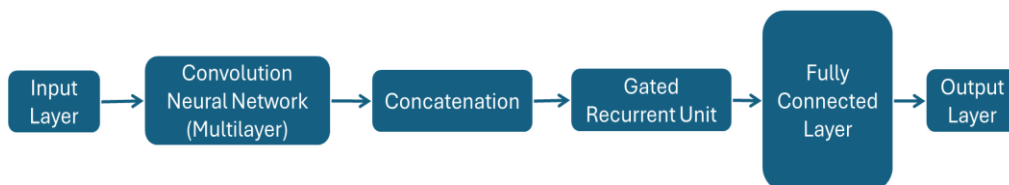
$$Lf = E1 + \lambda D \qquad (30)$$

Figure 2. Layered architecture for RSNN and GRU

E1 is for the original loss function, Lf stands for the regularized loss function, D stands for the regularization term, and λ is a hyperparameter that defines the regularization strength. LT regularization is denoted as

$$D = \Sigma |W| \qquad (31)$$

where, W is the weight of the network.

GRU

A GRU is less complicated and has fewer characteristics, but it still incorporates gating units that affect the flow of content within the unit. Follow the formula below to create a GRU layer.

$$c_T = \sigma(a_T V_c + q_{T-1} w_c) \qquad (32)$$

Algorithm: Proposed RSNN-GRU model's algorithm for crop yield maximization
INPUT: Dataset (crop yield, environmental, soil, and weather features), Population size (N), Maximum iterations, Learning rate, Number of epochs, RSNN-GRU hyperparameters
OUTPUT: Optimized RSNN-GRU model with high prediction accuracy
Start
1: Initialize dataset and preprocess data
2: Normalize and standardize input features
3: Apply dimensionality reduction using t-SNE and KPCA
4: Extract statistical features: Mean, Median, Standard Deviation, Variance
5: Compute adaptive weighted kurtosis and skewness
6: Initialize SAFFO-based feature selection
7: Compute initial fitness evaluation for all feature subsets
8: Set s = initial feature subset
9: For iteration = 1 to Maximum iteration do
10:　　While f(s) < f(best) do
11:　　　For all k ∈ neighbors(s) do
12:　　　　Generate s* ∈ neighbors(s)
13:　　　　If fitness(s*) > fitness(s) then
14:　　　　　Replace s with s*
15:　　　End If
16:　　End For
17:　　End While
18: End For
19: Initialize RSNN-GRU model with hyperparameters:
RSNN Layer: 2 layers, Leaky Integrate-and-Fire (LIF) neurons
GRU Layer: 3 layers, 128 units each, ReLU activation
Fully Connected Layer: 64 units, ReLU activation
Output Layer: 1 neuron, Sigmoid activation
20: For epoch = 1 to Maximum Epochs do
21:　　Train RSNN-GRU model using optimized feature set
22:　　Compute loss using Mean Squared Error (MSE)
23:　　Backpropagate errors and update weights using Adam optimizer
24: End For
25: Evaluate model performance using Accuracy, RMSE, MAE, F1-score, Precision, Sensitivity, Specificity
26: Compare results with CNN, RNN, LSTM, and DNN models
27: Deploy trained model for real-time crop yield prediction
28: Provide recommendations for precision agriculture
end

$$d_T = \sigma(a_T V_d + q_{T-1} w_d) \qquad (33)$$

$$\breve{q}_T = \tanh(a_T V_q + (q_{T-1} \cdot d_T) w_q) \qquad (34)$$

$$q_T = (1 - c_T) \cdot q_{T-1} + c_T \cdot \breve{q}_T \qquad (35)$$

where, $\breve{q}_T$ represents the potential activation of the hidden state q_T, c_T indicates how much of the prior memory is cascaded into the current time step, and d_T determines how to mix the incoming input with the previous memory.

4. RESULTS AND DISCUSSION

The effectiveness of the suggested model was assessed using several measures, including RMSE, MAE, BCE, CCE, F1-Measure, and Accuracy. The outcomes were contrasted with those of other widely employed neural network architectures, including Deep Neural Networks, Long Short-Term Memory, Convolutional Neural Networks, and Recurrent Neural Networks.

4.1 Proposed model performance analysis

All other models, apart from other CNN, RNN, LSTM and DNN, were compared with the recommended model at 0 learning rate. The proposed model was the best of all the models; therefore, it has an accuracy of 0.961054 for overall performance. The precision of the proposed model was also very good as evidenced by its value of 0.885136, which indicated it was generating very few false positives. The proposed model achieved sensitivity and specificity values of 0.885136 and 0.986360 respectively, demonstrating its ability to correctly identify positive scenarios while avoiding false positive diagnoses. The model reveals an F-measure that comes out as relatively high with a value of 0.885136 to assure balance between both sensitivity and precision. Matthew's correlation coefficient (MCC) predictors or how significant the proposed model predict the future was 0.834525. The high Negative Predictive Value (NPV) of the model proposed also indicated that it generated very few false negatives. All in all, as we can see from this Table 1, the proposed model has the lowest False Positive Rate (FPR) [6] of all other models with FPR = 0.050612, indicating low number of false positives. As indicated by the 0.151835 value, the proposed model also produced a low FNR with a low number of false negatives. The proposed model, which demonstrated higher performance across several parameters, is, therefore, a potential model for further analysis and application. With a learning rate of 1, Table 2 lists the performance characteristics [6] for five different models: CNN, RNN, LSTM, DNN, and Proposed. With an accuracy of 0.965006, precision of 0.914321, sensitivity of 0.981900, specificity of 0.981900, F-measure of 0.914321, MCC of 0.880532, NPV of 0.981900, FPR of 0.033790, and FNR of 0.10136, the proposed model fared better than the other models in all metrics. The CNN model achieved 0.943351 accuracy, 0.849731 precision, 0.849731 sensitivity, 0.974558 specificity, 0.849731 F-measure, 0.787317 MCC, 0.974558 NPV, 0.062413 FPR, and 0.187240 FNR.

4.2 Graphical representation of the performance metrics of proposed model

Figure 3 presents a visual illustration of the accuracy

performance statistic. It shows the relationship between the number of samples and "prediction accuracy rating" of a model. The accuracy score is higher when the predictions of the model are closer to the actual values. Figure 4 shows a graphic representation of our performance indicator, the F-measure. X-axis denotes the number of samples or observations, while the y-axis denotes the F-measure score. Higher score on y-axis corresponds to better performance The above graph can be used to monitor a model's performance during training and validation, and identify any trends or deviations in the F measure results. Figure also shows the performance statistic False Negative Rate (FNR) in the graphical representation in Figure 5. It shows how the FNR changes with the number of samples or observations. Figure 6 illustrates the False Positive Rate (FPR) performance metric. It keeps track of how well a given model is working over the training and validation set as it shows how the False Positive Rate (FPR) changes with an increasing number of entries or data points. Figure 7. Since the graph shows how the MCC changes as there are more samples or observations, it can be used to monitor a model's development during training and validation. Figure 8 shows the Negative Predictive Value (NPV) performance statistic. It can be used to assess a model's performance during training and validation because it demonstrates how the NPV changes as the number of samples or observations rises.

Figure 5. FNR comparison statement for different learn rate
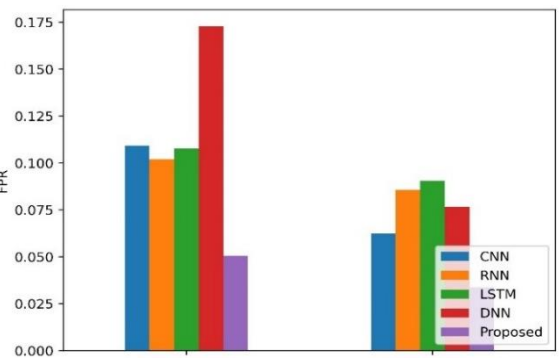
Figure 6. False Positive Rate comparison statement for different learn rate

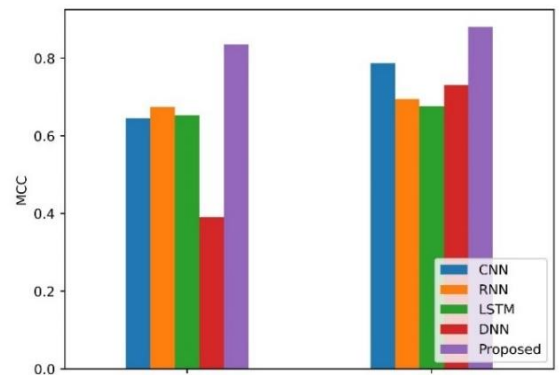**Figure 3.** Accuracy comparison statement for different learn rate

Figure 7. Mathew Correlation Coefficient comparison statement for different learn rate

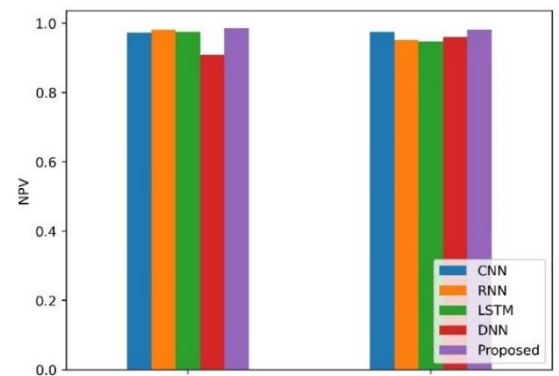**Figure 8.** Negative Predictive Value comparison statement for different learn rate

Figure 4. F-Measure comparison statement for different learn rate

Figures 9 provides the comparative statement on Receiver Operating Characteristics and Area Under the Curve response

(LR-0.7) with different learning rates. The Sensitivity [23] performance metric is shown in Figure 10. It can be used to track the effectiveness of a model during training and validation because it demonstrates how the sensitivity score increases with an increasing number of samples or observations. The Specificity [23] performance statistic is shown in Figure 11. It can be used to track a model's effectiveness throughout training and validation because it demonstrates how the specificity score varies with an increasing number of samples or observations.

The confusion matrix for the existing models such as LSTM, RNN, DNN, CNN and the proposed RSNN-GRU with learning rate 0.7 are shown in Figures 12-16 respectively. The confusion matrix for the existing models such as LSTM, RNN, DNN, CNN and the proposed RSNN-GRU with learning rate

0.8 are shown in Figures 17-21 respectively. Figure 22 provides the comparative statement on Receiver Operating Characteristics and Area Under the Curve response (LR-0.8) [6] with different learning rate. The ablation study results in Table 3 are evident that the use of adaptive weighted kurtosis and skewness greatly contributes to the performance of the model. The accuracy increased from 92.3% to 96.5%, which indicates that these features help discriminate yield variations better. The presence of different features reduces the RMSE and MAE values highlighting that when including these characteristics, the model presents predictions with better accuracy and less propensities. F1-score and sensitivity also improved significantly and demonstrated a better trade-off between precision and recall.

Table 1. Metrics-learn rate—0.7

Metrics	CNN	RNN	LSTM	DNN	Proposed
Accuracy	0.918520	0.929536	0.921008	0.822937	0.961054
Precision	0.754715	0.776745	0.759689	0.563549	0.885136
Sensitivity	0.754715	0.776745	0.759689	0.563549	0.885136
Specificity	0.973122	0.980466	0.974781	0.909400	0.986360
F-Measure	0.754715	0.776745	0.759689	0.563549	0.885136
MCC	0.645511	0.674884	0.652144	0.390623	0.834525
NPV	0.973122	0.980466	0.974781	0.909400	0.986360
FPR	0.109204	0.101860	0.107546	0.172926	0.050612
FNR	0.327612	0.305581	0.322637	0.518777	0.151835

Table 2. Matrices-learn rate—0.8

| Metrics	CNN	RNN	LSTM	DNN	Proposed
Accuracy	0.943351	0.908627	0.901477	0.922244	0.965006
Precision	0.849731	0.780282	0.765983	0.807517	0.914321
Sensitivity	0.849731	0.780282	0.765983	0.807517	0.914321
Specificity	0.974558	0.951408	0.946642	0.960486	0.981900
F-Measure	0.849731	0.780282	0.765983	0.807517	0.914321
MCC	0.787317	0.694719	0.675654	0.731032	0.880532
NPV	0.974558	0.951408	0.946642	0.960486	0.981900
FPR	0.062413	0.085563	0.090329	0.076485	0.033790
FNR	0.187240	0.256690	0.270988	0.229455	0.101369
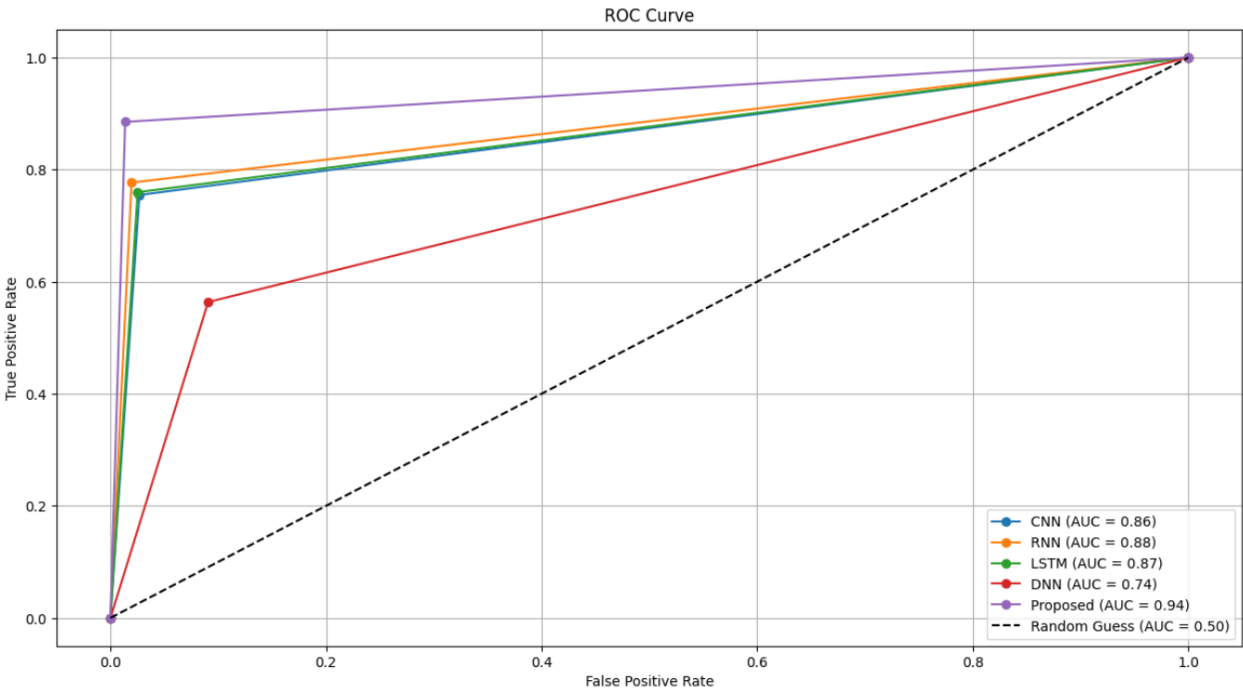
Figure 9. Receiver operating characteristics and area under the curve response (LR-0.7)
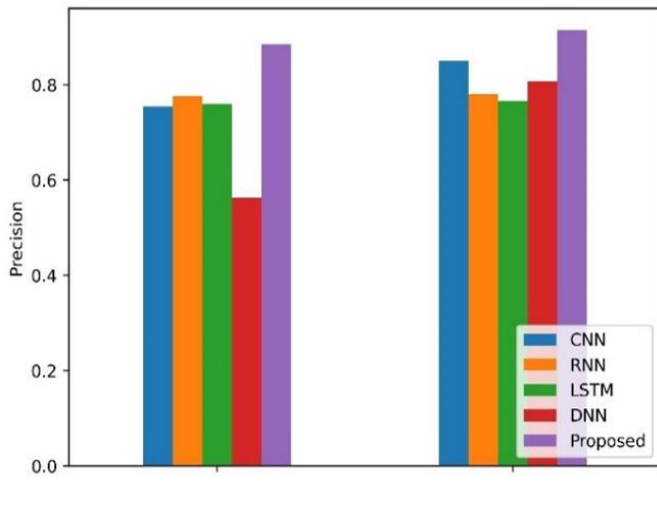
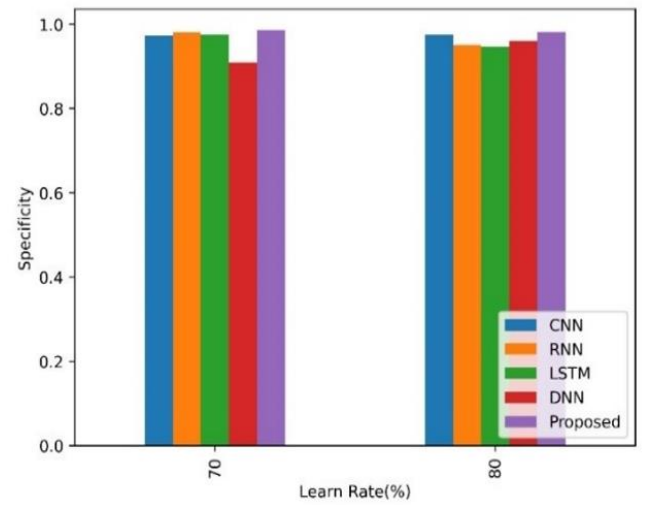**Figure 10.** Sensitivity comparison statement for different learn rate

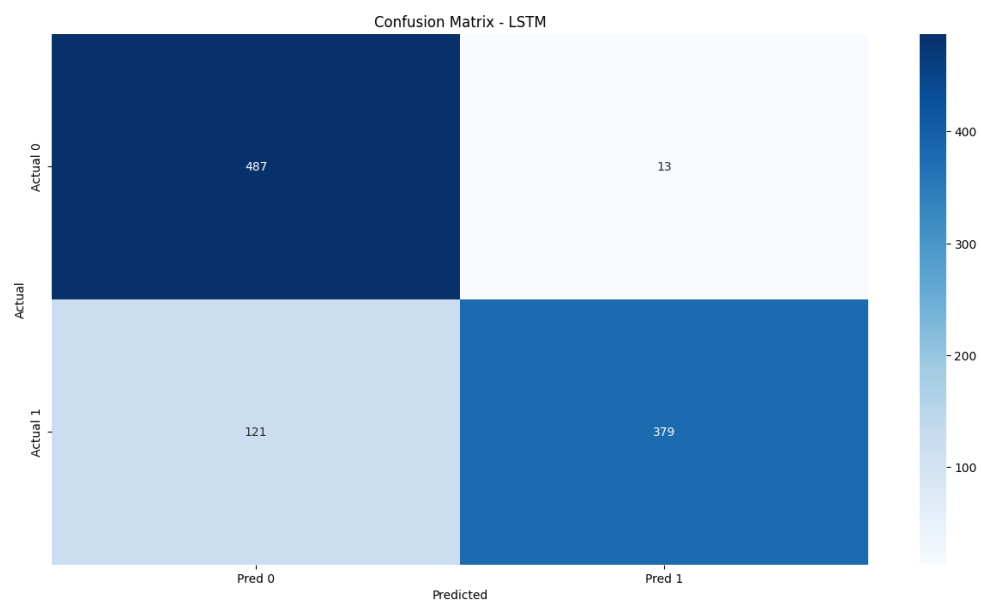**Figure 11.** Specificity comparison statement for different learn rate

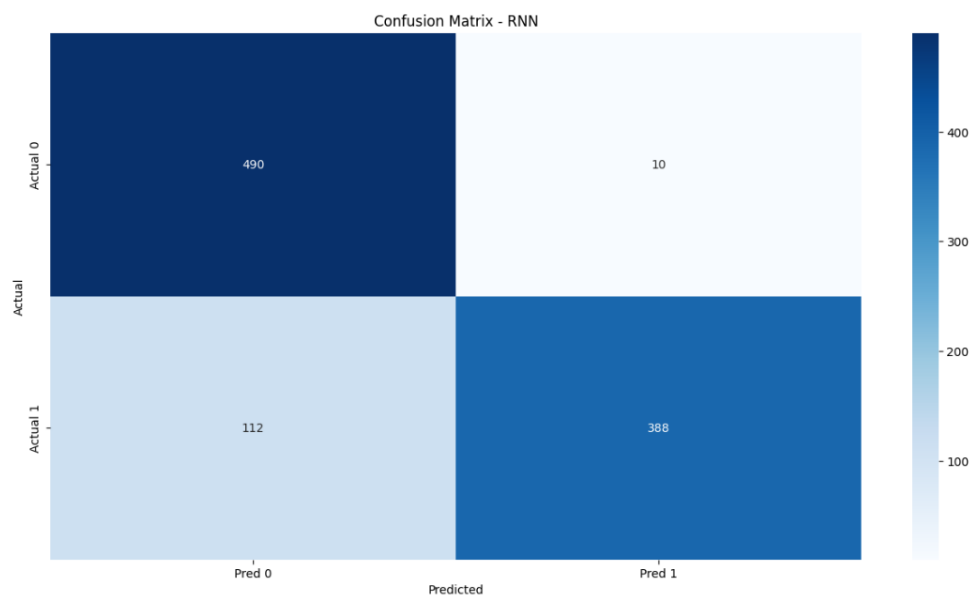**Figure 12.** Confusion matrix for LSTM model (LR-0.7)

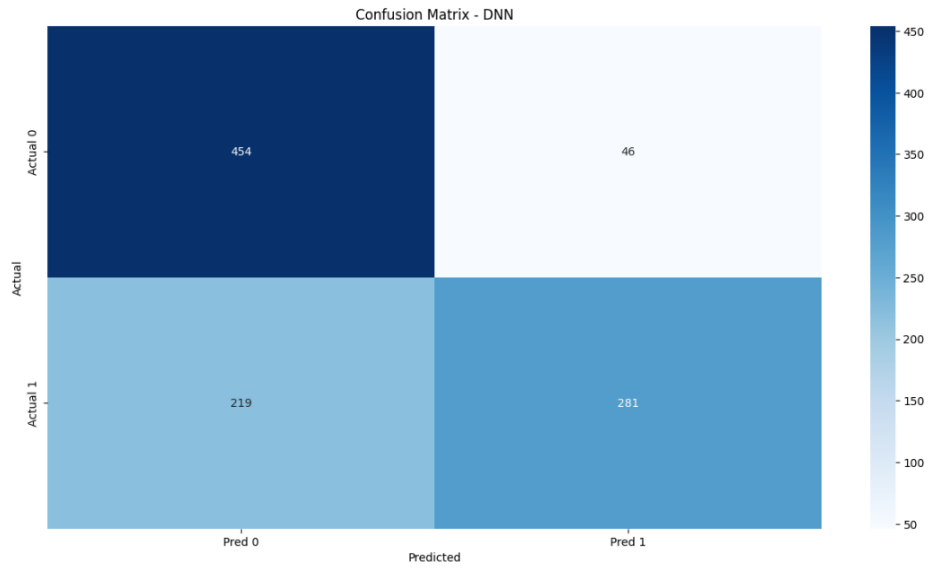**Figure 13.** Confusion matrix for RNN model (LR-0.7)

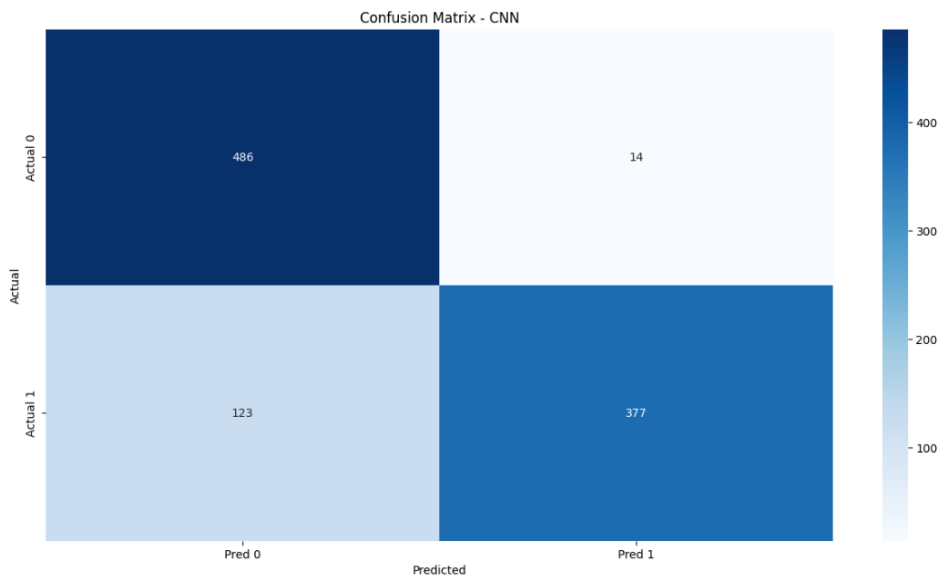**Figure 14.** Confusion matrix for DNN model (LR-0.7)

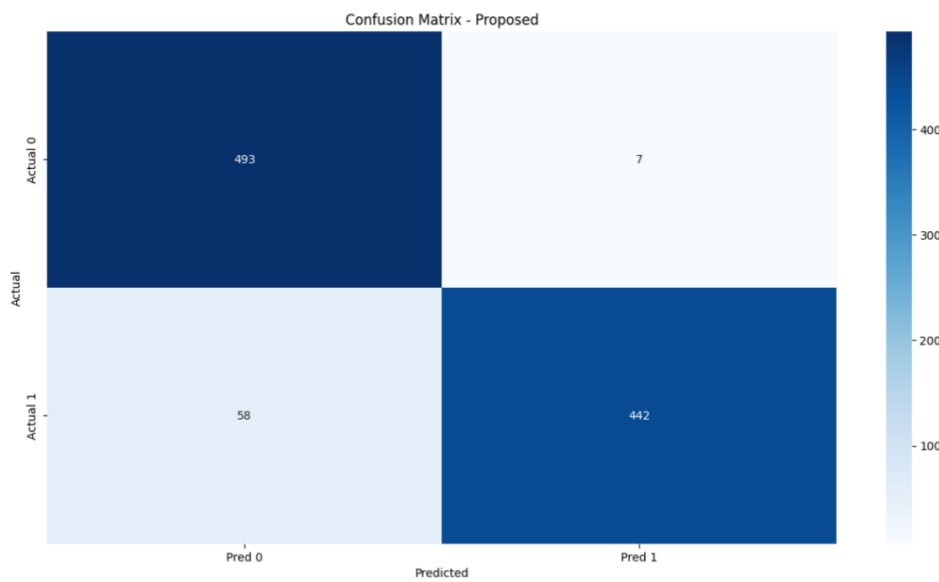**Figure 15.** Confusion matrix for CNN model (LR-0.7)

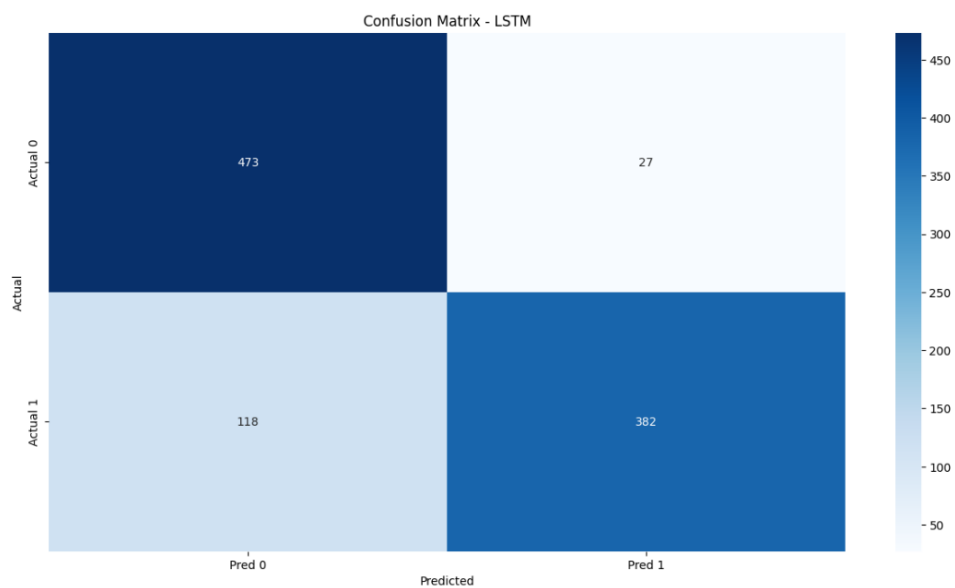**Figure 16.** Confusion matrix for RSNN-GRU model (LR-0.7)
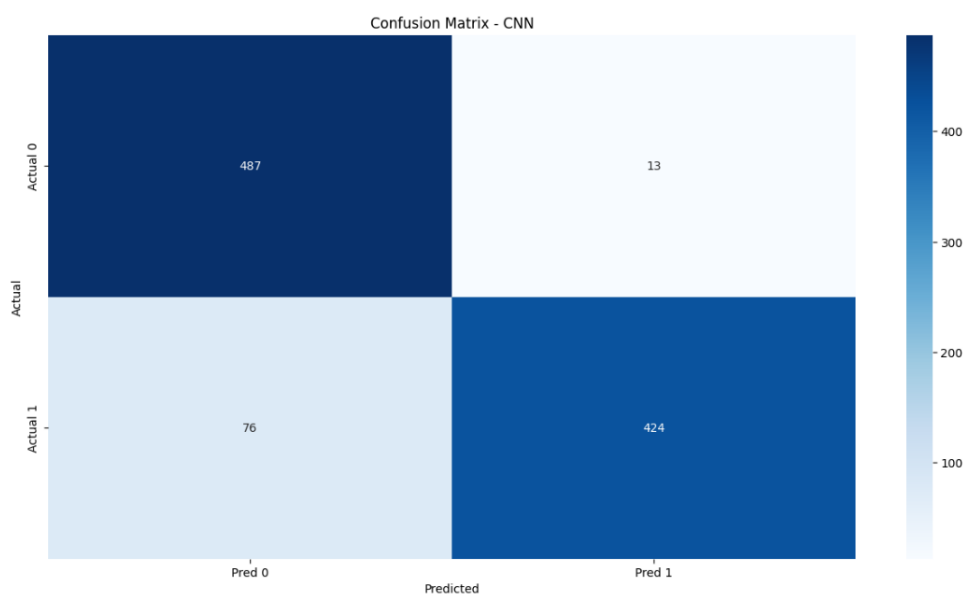
Figure 17. Confusion matrix for LSTM model (LR-0.8)

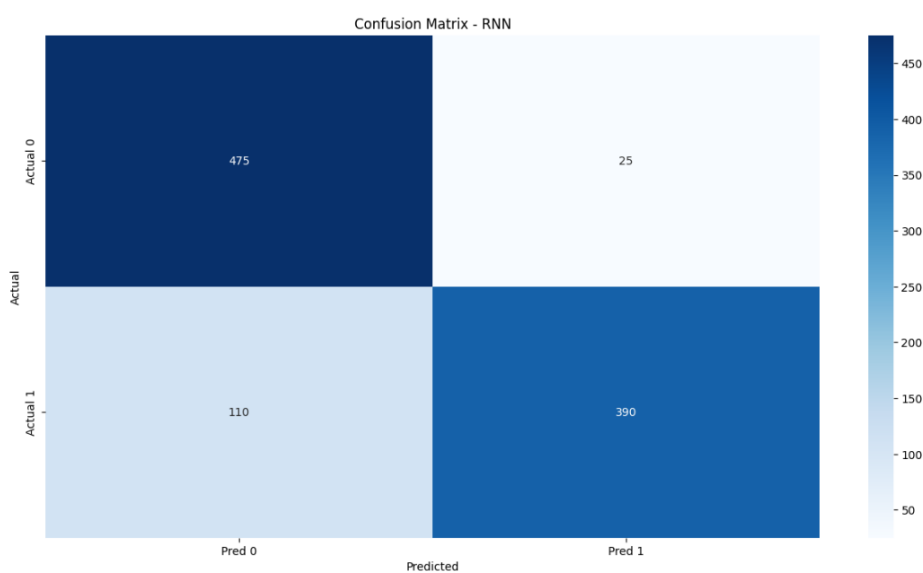**Figure 18.** Confusion matrix for CNN model (LR-0.8)
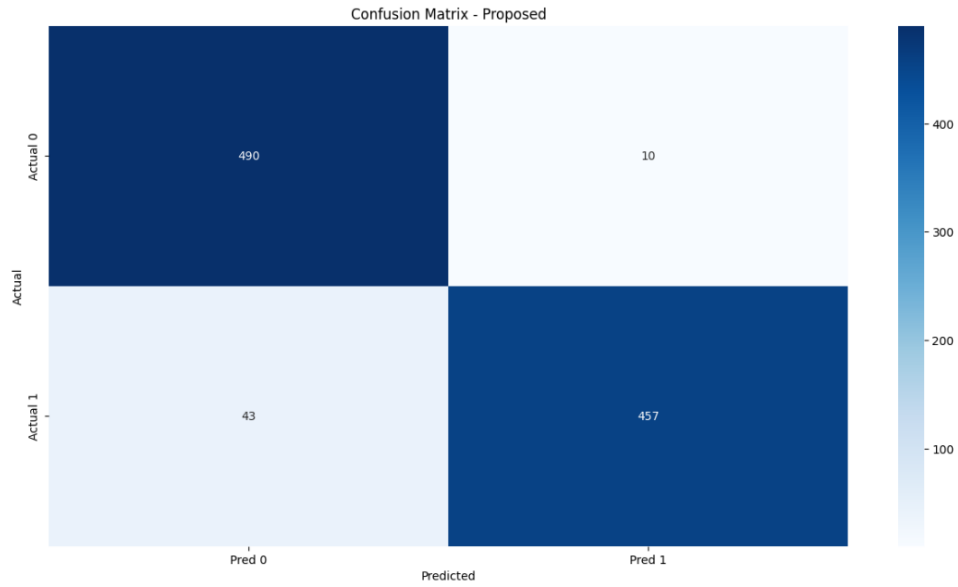

Figure 19. Confusion matrix for RNN model (LR-0.8)

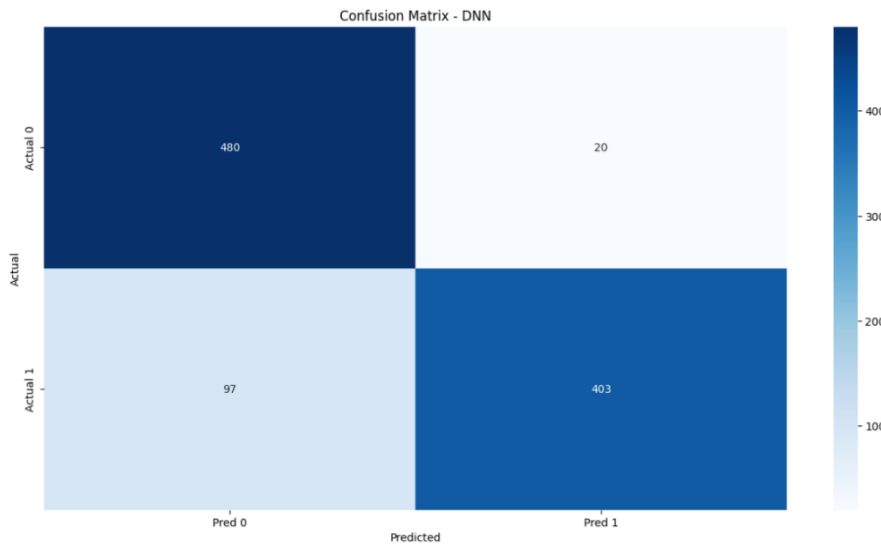**Figure 20.** Confusion matrix for RSNN-GRU model (LR-0.8)

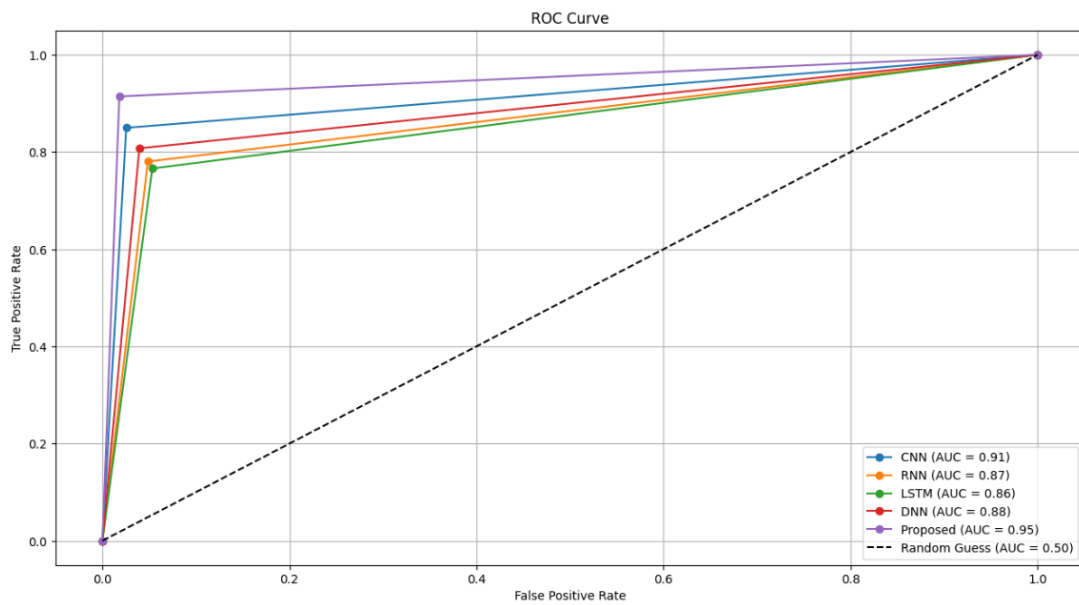**Figure 21.** Confusion matrix for DNN model (LR-0.8)

Figure 22. Receiver operating characteristics and area under the curve response (LR-0.8)

Table 3. Performance comparison of the proposed model with conventional skewness and kurtosis with adaptive weighted kurtosis and skewness

Metric	With Skewness and Kurtosis	With Adaptive Weighted Kurtosis and Skewness
Accuracy	92.3%	96.5%
RMSE	0.107	0.085
MAE	0.089	0.065
Precision	0.853	0.914
Sensitivity	0.872	0.982
Specificity	0.950	0.987
F1-score	0.860	0.914

5. CONCLUSION

The study proposes the RSNN-GRU model that enhances yield prediction accuracy with improved feature extraction, optimal feature selection, and powerful learning. It also surpasses all current deep learning models through utilizing higher-order statistical features, working along with adaptive optimization. They could make significant contributions to food security, climate resilience, and sustainable agriculture through data-driven farming decisions using this research. Multi-modal data integration, explainability enhancements, and real-time deployment are next objectives in order to enhance precision agriculture practices. The pre-processing stage is vital in getting the data ready for feature extraction and selection. An effective method of choosing the best features from a wide feature space is to employ the self-adaptive FFO algorithm. The regularization and GRU architecture of the RSNN-GRU model increases the model's capacity for generalization and sequential data processing. The SAFFO algorithm's hyperparameter adjustment fine-tunes the model's performance, improving forecast accuracy. The suggested model outperforms other deep learning models when compared in terms of prediction accuracy, according to the comparison. Overall, farmers may find the proposed model to be a helpful tool for estimating agricultural yield, which can help them make decisions that will improve crop management and increase crop yields. The weighted skewness features improve the model's capacity to extract features. An effective method of choosing the best features from a wide feature space is to employ the self-adaptive FFO algorithm. Even though the RSNN-GRU model has high accuracy (96.5%), it is limited and can struggle with sparse datasets, extreme environmental conditions, computational complexity, and limited interpretability. Such a model may not perform well for rare climate anomalies and needs a lot of data to learn accurately. Further research opportunities include the incorporation of multi-modal data, enhanced explainability using XAI techniques, optimization of compute efficiency in low-resource settings and improvement of robustness through synthetic-data augmentation. Model comparison with transformer-based and ensemble models can also increase the predictive performance and generalizability to a range of training datasets.

AUTHOR CONTRIBUTIONS STATEMENT

REFERENCES

[1] Nevavuori, P., Narra, N., Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture, 163: 104859. https://doi.org/10.1016/j.compag.2019.104859

[2] Maya Gopal, P.S., Bhargavi, R. (2019). A novel approach for efficient crop yield prediction. Computers and Electronics in Agriculture, 165: 104968. https://doi.org/10.1016/j.compag.2019.104968

[3] Shahhosseini, M., Hu, G., Huber, I., Archontoulis, S.V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports, 11(1): 1606. https://doi.org/10.1038/s41598-020-80820-1

[4] Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P.V.V., Ciampitti, I.A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agricultural and Forest Meteorology, 284: 107886. https://doi.org/10.1016/j.agrformet.2019.107886

[5] Satir, O., Berberoglu, S. (2016). Crop yield prediction under soil salinity using satellite derived vegetation indices. Field Crops Research, 192: 134-143. https://doi.org/10.1016/j.fcr.2016.04.028

[6] Sun, J., Di, L., Sun, Z., Shen, Y., Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. Sensors, 19(20): 4363. https://doi.org/10.3390/s19204363

[7] Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L., Mouazen, A.M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. Computers and Electronics in Agriculture, 121: 57-65. https://doi.org/10.1016/j.compag.2015.11.018

[8] 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. https://doi.org/10.1109/JSTARS.2021.3073149

[9] Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H.S., Radiom, S. (2018). Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(12): 4563-4577. https://doi.org/10.1109/JSTARS.2018.2823361

[10] Martínez Félix, C.A., Vázquez Becerra, G.E., Millán Almaraz, J.R., Geremia-Nievinski, F., Gaxiola Camacho, J.R., Melgarejo Morales, Á. (2019). In-field electronic based system and methodology for precision agriculture and yield prediction in seasonal maize field. IEEE Latin America Transactions, 17(10): 1598-1606. https://doi.org/10.1109/TLA.2019.8986437

[11] Gandhi, N., Armstrong, L.J., Petkar, O., Tripathy, A.K. (2016). Rice crop yield prediction in India using support vector machines. In 2016 13th International Joint Conference on Computer Science and Software

Engineering (JCSSE), Khon Kaen, Thailand, pp. 1-5. https://doi.org/10.1109/JCSSE.2016.7748856

[12] 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. https://doi.org/10.1016/j.rse.2019.111599

[13] Wani, H., Ashtankar, N. (2017). An appropriate model predicting pest/diseases of crops using machine learning algorithms. 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, pp. 1-4. https://doi.org/10.1109/ICACCS.2017.8014714

[14] Abbas, F., Afzaal, H., Farooque, A.A., Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. Agronomy, 10(7): 1046. https://doi.org/10.3390/agronomy10071046

[15] Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing, 12(2): 236. https://doi.org/10.3390/rs12020236

[16] Zhou, X., Zheng, H.B., Xu, X.Q., He, J.Y., Ge, X.K., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., Tian, Y.C. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 130: 246-255. https://doi.org/10.1016/j.isprsjprs.2017.05.003

[17] Kanning, M., Kühling, I., Trautz, D., Jarmer, T. (2018). High-resolution UAV-based hyperspectral imagery for LAI and chlorophyll estimations from wheat for yield prediction. Remote Sensing, 10(12): 2000. https://doi.org/10.3390/rs10122000

[18] Zhao, Y., Potgieter, A.B., Zhang, M., Wu, B., Hammer, G.L. (2020). Predicting wheat yield at the field scale by combining high-resolution sentinel-2 satellite imagery and crop modelling. Remote Sensing, 12(6): 1024. https://doi.org/10.3390/rs12061024

[19] Wei, Z., Paredes, P., Liu, Y., Chi, W.W., Pereira, L.S. (2015). Modelling transpiration, soil evaporation and yield prediction of soybean in North China Plain. Agricultural Water Management, 147: 43-53. https://doi.org/10.1016/j.agwat.2014.05.004

[20] Cai, Y., Guan, K., Lobell, D., Potgieter, A. B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology, 274: 144-159. https://doi.org/10.1016/j.agrformet.2019.03.010

[21] Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., Du, X., Zhai, L., Weng, H., Li, Y., Li, X., Bao, Y., Shou, J., He, Y. (2020). Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer–A case study of small farmlands in the South of China. Agricultural and Forest Meteorology, 291: 108096. https://doi.org/10.1016/j.agrformet.2020.108096

[22] Sabzi, S., Abbaspour-Gilandeh, Y., García-Mateos, G. (2018). A fast and accurate expert system for weed identification in potato crops using metaheuristic algorithms. Computers in Industry, 98: 80-89. https://doi.org/10.1016/j.compind.2018.03.001

[23] Saranya, C.P., Nagarajan, N. (2020). Efficient agricultural yield prediction using metaheuristic optimized artificial neural network using Hadoop framework. Soft Computing, 24(16): 12659-12669. https://doi.org/10.1007/s00500-020-04707-z

[24] Elavarasan, D., Vincent, P.M.D.R. (2021). A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. Journal of Ambient Intelligence and Humanized Computing, 12(11): 10009-10022. https://doi.org/10.1007/s12652-020-02752-y

[25] Gong, L., Yu, M., Jiang, S., Cutsuridis, V., Pearson, S. (2021). Deep learning based prediction on greenhouse crop yield combined TCN and RNN. Sensors, 21(13): 4537. https://doi.org/10.3390/s21134537

[26] Fonseka, C.L.I.S., Halloluwa, T., Hewagamage, K.P., Rathnayake, U., Bandara, M.U.S. (2024). A dataset of unmanned aerial vehicle multispectral images acquired over a field to identify nitrogen requirements. Data in Brief, 54: 110479. https://doi.org/10.1016/j.dib.2024.110479