IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 September 2024, accepted 14 October 2024, date of publication 5 November 2024, date of current version 13 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3486101

RESEARCH ARTICLE

Enhancing Intercropping Yield Predictability Using Optimally Driven Feedback Neural Network and Loss Functions

AMNA IKRAM[1] AND WAQAR ASLAM[2]
[1]Department of Computer Science and IT, Government Sadiq College Women University, Bahawalpur 63100, Pakistan
[2]Department of Information Security, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

Corresponding author: Waqar Aslam (waqar.aslam@iub.edu.pk)

ABSTRACT Enhancing the crop yield predictability in intercropping systems is important for optimizing agricultural productivity. However, accurately predicting yield in such systems is quite challenging due to complex interactions between crops. This study introduces an advanced methodology using integrated loss functions within an optimally driven Feedback Neural Network (FNN) approach to improve yield prediction in a pea-cucumber intercropping systems. Traditional models relying only on Mean Square Error (MSE) loss function often unable to capture the complexity of models, leading to suboptimal performance. To address this limitation, the advanced loss functions are introduced like Dynamic Margin Loss (DML), Risk-Adjusted Loss (RAL), Quantile Loss (QL), and Hybrid Agronomic Efficiency Loss (HAEL) along with three optimizers such as Adaptive Momentum (Adam), Root Mean Square Propagation (RMSprop), and Adaptive delta (Adadelta). These loss functions incorporate risk, uncertainty, and agronomic efficiency into the model training process, enhances predictive capabilities and robustness. This proposed framework is able to capture the complexity of yield prediction by incorporating agricultural factors. While Gradient Boost Machines (GBM) and Long Short Term Memory (LSTM) have some potential, they are not able to capture these dynamics. The sensitivity and weight analysis also focuses that HAEL targets important agronomic factors such as nitrogen uptake and residue biomass, which provide a holistic view of yield prediction. The proposed approach improves the predictive performance compared to traditional models and helps to identify the importance of features, which makes it an effective tool for decision making in sustainable agriculture. Selecting appropriate loss functions is essential to improve the accuracy and robustness of crops yield prediction models. Thus, study provides a strong foundation for enhancing yield prediction in intricate intercropping systems, which all significantly enhance the advancement of precision agriculture.

INDEX TERMS Smart agriculture, pea-cucumber intercropping, yield prediction, artificial neural network, long short term memory, loss functions.

I. INTRODUCTION

Intercropping is a sustainable farming technique which involves cultivating multiple crops together in the same field, offering advantages like yield maximization, effective utilization of resources and improving biodiversity [1]. However, predicting crop yield in intercropping systems can be quite challenging due to complex interactions between crops [2], [3], [4]. The traditional models using MSE may inadequately address agricultural realities, being sensitive to outliers and ignoring practical implications of errors [5], [6]. Depending on MSE for yield prediction can give inaccurate results because it doesn't capture the complexities of agricultural systems [7]. To address these problems, the study utilizes multiple advanced loss functions in an optimized Feedback Neural Network framework (FNN) which we named in our case as FNNIS. These loss functions, DML, RAL, QL, and HAEL aim to enhance prediction accuracy by incorporating risk, uncertainty, and practical relevance considerations.

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Amitrano.

DML learns the margin based on real-time learning of error estimates of the prediction errors that are derived from the difference between the predicted and actual values and adapted its margin accordingly [8]. This flexibility is important in diverse agricultural systems that are characterized by factors such as weather, soil type and crops interrelationships that cause large deviations in the predictions. In this way, DML is able to maintain models' robustness and their ability to handle the volatility, typical of changing agricultural environments.

RAL assess the uncertainty in yield predictions by providing information about the probability of achieving desired results under different conditions. This enables the farmers and decision makers to take decisions at different risk levels, and assists them especially for better and more sustainable planning [7].

QL is an improvement over simple yield prediction since it provides yield predictions at different quantiles of the yield distribution. This is especially useful for managing risks with extreme yield volatility in agriculture systems. QL also assists in anticipating both the worst and best case scenarios because it gives broader view of probable consequences [9].

HAEL combines traditional MSE loss with components that focus on agronomic efficiency and yield discrepancy. This integration optimizes both prediction accuracy and practical utility for agriculture. By balancing prediction errors with real world agricultural applications, HAEL ensures that the model's predictions are not only accurate but also highly relevant and useful to every day farming practices. This makes HAEL particularly effective in refining yield predictions, ensuring that the model's output are both accurate and practically applicable.

Evaluating these loss functions allow us to address various aspects of prediction challenges comprehensively [10], [11]. This approach has made our model more robust and sustainable in agricultural context, demonstrating its superiority over traditional MSE-based models and machine learning models based on different architectures in optimizing yield predictions. Optimizers such as Adam, RMSprop, and Adadelta impact the effectiveness of these loss functions. Aligning loss functions with agricultural needs improves yield prediction accuracy, supporting better decision making and promoting sustainable intercropping practices. By employing advanced techniques, farming efficiency and sustainability can be enhanced.

For proposed study, the research questions are as follows:

RQ1: How is the performance of a Feedback Neural Network (FNN) in predicting crops yield in an intercropping agricultural system? FNN in our case is termed as FNNIS and it uses a loss function to minimize prediction errors.

RQ2: How is the performance of various loss functions in FNNIS? The tested loss functions are Dynamic Margin Loss (DML), Risk-Adjustment Loss (RAL), Quantile Loss (QL), and Hybrid Agronomic Efficiency Loss (HAEL).

RQ3: How is the performance of various optimizers in FNNIS? The tested optimizers are Adam, RMSprop and Adadelta.

Intercropping in agricultural crops is an emerging trend, offering promising results towards efficient utilization of resources and increased productivity. Various agricultural approaches are increasingly being adopted for yield prediction. To this end, studying the applicability of Artificial Intelligence (AI) techniques in this domain is appealing, though the new opportunities also bring in unseen challenges. The main objective of this work is to enhance the crops yield prediction accuracy and reliability in an intercropping agriculture system using AI machinery such as Neural Networks and optimization algorithms.

We contributed by proposing a Feedback Neural Network in an intercropping agricultural system for predicting the crops yield with high accuracy. Various loss functions and optimizers are tested with this Feedback Neural Network for exploring results beyond benchmark Neural Networks.

In Section I, the Introduction of the work is presented. Section II provides a literature review, contextualizing the study within the related existing research. Section III describes the methodology of the proposed model and features the dataset. Section IV presents the results and analysis. Section V provides the discussion about results and analysis. Finally, Section VI, conclusion summarizes the findings and give limitations of study.

II. RELATED WORK

This section comprises existing literature related to proposed approach for yield prediction in intercropping systems using deep learning approaches. The related work is organized into several key areas:

A. YIELD PREDICTION MODELS IN INTERCROPPING SYSTEMS

The increase in data availability across various scales, ranging from country-wide to field-level, has opened the door for new data-centric methodologies. These process-based theories are integrated into what are known as "crop growth model," which, in essence, encapsulate physical processes. The combination of plant species with different traits can improve resource utilization and improve overall yield in intercropping systems [12]. The crop type and sowing timing influence the biomass residue, nutrient absorption, and improve intercrop production in intercropping systems [13]. The cucumber with green garlic can improve soil and plant nutrition by highlighting the significance of crop interactions in low nitrogen systems [14]. Furthermore, cultivation of maize in intercropping systems demonstrated high yield benefits attributed to the improvement of Leaf Area Index (LAI), allocation of dry matter, leading to higher chlorophyll levels and specific leave weight (ratio of leave dry weight to leaf area) [15].

Peanut-cotton intercropping has been demonstrated to improve crop production and return on investment across diverse soil conditions categorized as saline, secondary-saline, coastal saline and non-saline. This improvement is most likely due to changes in soil bacterial quantity and composition, as well as a faster buildup of plant nutrients [16]. These findings provide a better knowledge of the underlying processes that govern and characterize the yield benefits and profit returns in intercropping systems [17]. The recent advancements in deep learning domain have significantly influenced agricultural predictions, primarily depending on data driven approaches for yield forecasting [18], [19], [20]. Machine learning algorithms are highly beneficial to improve agricultural production efficiency, assisting in decision-making related to planting, watering, harvesting and grading the crops [21], [22]. A deep learning framework that combines Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and LSTM has been implemented for predicting crop yields [23]. The artificial neural network (ANN) in crop yield prediction has proven to be highly effective across various studies, outperforming traditional models like Multiple Linear Regression (MLR) by accurately capturing nonlinear relationships between yield and agronomic traits, with applications in rice, wheat, apple and hybrid MLR-ANN models showing significant improvements in prediction accuracy and computational efficiency [24], [25], [26], [27], [28]. The implementation of push-pull technology in maize-legume intercropping aims to control pest and disease factors affecting crop yield, achieved by using hybrid fuzzy logic models coupled with a genetic algorithm to predict maize yield [18]. The linear algorithms Linear Discriminant Analysis (LDA) and Logistic Regression (LR) demonstrate better alignment between predicted and actual maize yields than non-linear methods such as Naïve Bayes (NB), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), and Support Vector Regressor (SVM) [29].

B. ADVANCED DEEP LEARNING APPROACHES FOR ENHANCING PREDICTING CAPABILITY

Advanced modeling approaches play an important role in enhancing yield prediction within intercropping systems. These approaches use deep learning methodologies and other efficient algorithms to capture complex interactions and improve predictive accuracy [30]. The decision based models play an important role in minimizing risks associated with decision making and ensuring sound judgments. Therefore, the majority of prediction models has limited performance and cannot handle complex problems. The Enhanced Conceptual Pre-prediction (ECP) model, which utilize three-way decision (TWD) concepts, have been developed to improve prediction accuracy [31]. In the context of energy storage system arbitrage, decision focused electricity price prediction emphasized the influences of predictions on downstream decision making, rather than just minimizing prediction errors [32]. Recent advancement in data assimilation and weak-constraint 4D-Var have shown promising results in correcting systematic model errors in numerical weather prediction and climate forecasting [33]. Using machine learning models and ensembling them for crop yield prediction in the US Corn Belt, demonstrates that a decision based model combining crop modeling significantly enhances accuracy [34].

C. OPTIMIZERS FOR ACCURACY ENHANCEMENT

Optimization algorithms enhance model accuracy by refining learning process through repeated cycles until convergence is achieved. Various optimization techniques have been developed to address these challenges encountered during the learning process [35]. Recent advancements in optimizer selection, such as Adam optimizer, have shown notable improvements in accuracy, with studies highlighting its ability to effectively handle both underfitting and overfitting challenges, thereby outperforming traditional methods in predictive tasks [36]. A comprehensive comparison of various optimizers for CNN-based tumor segmentation on the BraTS2015 dataset demonstrated that Adam optimizer achieved higher accuracy of 99.2%, underscoring its effectiveness in improving classification and segmentation performance [37]. The selection of optimizers such as Adam, which achieved 96% accuracy using CNN models, highlights its effectiveness in improving model performance when compared to other optimizers like RMSprop and SGD [38]. The Adam optimizer has been widely adopted due to its adaptive learning rate mechanism, leading to improved accuracy in crop yield prediction models, as demonstrated in the context of Andhra Pradesh for predicting rice, sugarcane, and onion yields [39].

D. ADVANCE LOSS FUNCTIONS

In recent studies, there has been developing interest to design custom loss functions for applications, aiming to enhance model performance and address specific challenges. A summary is given in Table 1. Not all prediction errors in yield prediction impact algorithm's effectiveness equally. The small errors can lead to suboptimal agricultural management decisions, effecting yield outcomes. On the other hand, larger errors might not significantly impact the final yield or agricultural practices. The data balance in medical image segmentation is performed with a generalized focal loss function based on Tversky index. It achieves a better balance between precision and recall, particularly for small structures like lesions, and enhances segmentation accuracy across diverse datasets [40]. Economic imbalance is important for assessing algorithm performance and can be factored into deep learning algorithms through the loss function. For instance, an efficient custom loss function has been designed that significantly enhances asset return predictions, improving risk-return metrics like Sharpe ratio. The results confirm its robustness [41]. The loss function is commonly used in regression and cross-validation when the true distribution is unknown, often results in overfitting. A modified L2-adjusted check loss is used that smooth the sharp corner in the check loss, reducing the risk of over fitting. This adjustment makes the loss function less sensitive to extreme values, making model more generalized [42]. A margin-aware adaptive-weighted loss function is introduced for class imbalance problems by using large margin softmax for class separation and dynamically adjusting weights based on inverse class frequencies and confidence scores [43]. An improved triplet loss function is proposed for end to end metric learning with deep neural networks, which reduced the number of possible triplets without compromising model performance. The model outperforms the previous approaches by achieving efficient results [44]. Additionally, smart "predict, then optimize" (SPO) loss has been suggested for handling optimization problems by learning linear predictor parameters using linear programming [45]. An enhanced deep learning algorithm with novel combined loss function (binary focal loss and dice loss) is proposed for skin lesion segmentation. This approach achieves state of art accuracy on dataset, demonstrating significant advancements in semantic segmentation [46]. Balanced MSE as a solution is proposed to improve regression tasks with imbalanced label distributions. It adapts multiple implementations to handle diverse real-world scenarios, demonstrating its effectiveness in high-dimensional imbalanced regression tasks [47].

TABLE 1. Advancements in loss functions.

Citations	Application context	Loss Function type	Key advancement
[40]	Medical Image Segmentation	Generalized Focal Loss	Better precision-recall balance, enhanced segmentation accuracy.
[41]	Financial Prediction	Custom Loss	Enhances asset return predictions, improves Sharp ratio metrics.
[42]	Regression-Cross Validation	L2-Adjusted Check Loss	Reduces overfitting, enhances model generalization
[43]	Class Imbalance	Adaptive-weighted Loss	Uses large margin softmax for class separation, dynamically adjusts weights based on inverse class frequencies and confidence score.
[44]	Metric Learning	Improved Triplet Loss	Reduces potential triplets, enhances efficiency in metric learning tasks.
[45]	Optimization problems	SPO Loss	Learns linear predictor parameters using linear programming, optimizing predictive accuracy and efficiency in optimization tasks.
[46]	Skin Lesion segmentation	Combined Binary Focal and Dice Loss	Improved accuracy, semantic segmentation
[47]	Regression	Balanced MSE	Accommodated imbalanced label distribution, improved accuracy

III. MATERIAL AND METHODS

The proposed approach aims to assess the interpretability and precision of deep learning techniques in the context of predicting yield in intercropping scenarios. Neural network models not only produce accurate results but also demonstrate resilience techniques for re-predicting the prediction error in each region, allowing the NN model's predicted values to be correct [39].

We divided FNNIS into two modules; one for data preprocessing and other for training module which we showed in Figure 1.

A. DATA PREPARATION MODULE

1) DATA COLLECTION

A research on pea-cucumber intercropping was carried out across 14 farms located in the eastern coastal highlands of Santa Catarina, spanning from May 2018 to December 2019 [48]. Dataset was acquired from a public repository at the following link:

https://zenodo.org/records/6079944

(Zenodo - pea-cucumber Intercropping Dataset) [49].

Initially, the dataset contained fifty variables with missing values. We subsequently removed few variables that were related to previous study [48], and were irrelevant to proposed model. Additionally, we derived new variables such as soil_organic_carbon, N_P ratio, nutrient_balance and total_biomass based on the available information within the dataset.

2) DATASET DESCRIPTION

The dataset contains 45 features and 225 records. The target variable Y is the yield. The ANN model takes the dataset as input and learns to predict the yield $\hat{Y}$ based on the input features.

3) FEATURES SELECTION

The deep learning models, especially neural networks, designed to automatically learn hierarchical representations from input data; but they may still benefit from feature selection techniques in certain scenarios. Traditional machine learning approaches often rely heavily on explicit feature selection, whereas deep learning models inherently learn representations from the data [50], [51]. However, in cases involving high-dimensional data or where interpretability is crucial, feature selection remains relevant even in deep learning models. For FNNIS, we utilized multiple approaches for feature importance assessment and selection. Expert opinion provided domain-specific insights, and selection of relevant features. Additionally, Principal Component Analysis (PCA) is also used to find the variability of data and most informative features while discarding irrelevant and redundant features. This combined strategy helped to simplify the input space for deep learning model, improving its efficiency and interpretability.

4) STANDARDIZATION

Standardization is a common preprocessing step for data used in deep learning algorithms, which is also known as Z-score normalization. Standardizing means the data is transferred to have zero mean and unit variance [52]. This process is important for deep learning algorithms where it helps in ensuring that the features are on a similar scale. It improves the training stability and convergence of model. The steps for standardizing the data are given next:

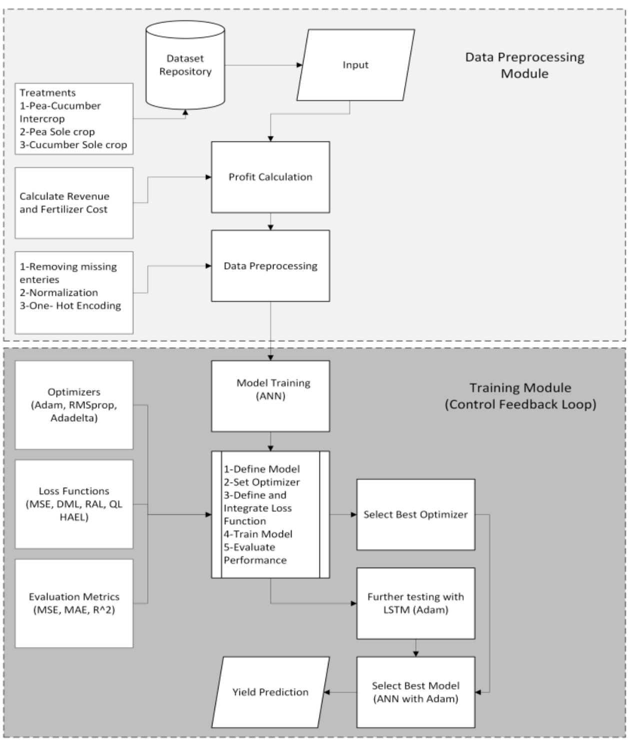

FIGURE 1. Workflow diagram of FNNIS for yield prediction.

For each parameter in dataset, the mean and standard deviation is calculated. Standardize each parameter (x) using the formula:

$$Z = \frac{x - \mu}{\sigma}, \tag{1}$$

where

- Z is the standardized value
- x is an input data
- μ is the mean of dataset
- σ denotes the standard deviation of dataset

Standardizing data is a critical step for deep learning algorithms. It transforms all the features on a consistent scale, promoting effective training of neural network models.

5) DATA ENCODING

A one hot encoding technique was applied to non-numeric columns ‘crop’ and ‘treatment’ in the dataset, converting them into binary vectors necessary for input into neural network models, specifically for further processing in FNNIS.

B. TRAINING MODULE

1) MODEL ARCHITECTURE (ANN)

In this study, we employed neural network (ANN) for yield prediction due to their ability to model complex, non-linear relationships between input features and target variable. Neural networks excel in capturing intricate patterns and interactions within data, which is essential for accurately predicting agricultural yield, characterized by multiple influencing factors and complex relationships in intercropping dynamics. The architecture of artificial neural network (ANN) is based on concept of human nervous system, which is widely used across various applications due to its ability to dynamically learn from data, extract patterns, and create frameworks for tasks like classification and regression [53]. Neural network models have ability to learn structured networks by adjusting parameters based on the contribution to computational loss. The ANN model typically comprises layers of processing nodes, each operating in parallel and arranged in sequential layers. Input data is initially received by the first layer, same as the function of optic nerves in the human visual system. Successive layers process the output from the previous layer using weights, biases, and activation functions, similar to how neurons interact in the nervous system [54]. The final output is produced by the last layer. ANN’s key feature is its self-adaptive nature, allowing it to adjust weights and biases iteratively based on their impact on computational loss, thereby optimizing performance across different datasets [55].

2) LOSS FUNCTIONS

To enhance the predictive accuracy and robustness of FNNIS model, we integrated various loss functions with MSE loss during training phase. These loss functions are designed to improve prediction accuracy and enhance the effectiveness of decision making processes, for the unique characteristics of our specific domain [56].

a: MEAN SQUARED ERROR (MSE)

The MSE loss is used during model training to optimize parameters especially in regression models where the independent variable is continuous. It is computed by taking average square of difference between predicted $\hat{Y}$ and the actual values Y for a dataset with N observations given in (2) [57].

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{Y}_i - Y_i\right)^2 \quad (2)$$

b: DYNAMIC MARGIN LOSS (DML)

The DML function improves the quality of the prediction by solving two major problems in regression tasks, namely underfitting and overfitting of the model. Unlike standard loss functions that do not distinguish between small and large errors, which can lead to suboptimal performance, particularly in datasets, where extreme deviations from the mean impact decision making process.

DML enhances prediction accuracy by introducing threshold-based penalization scheme through Marginal Penalty (MP) component.

i) DYNAMIC PENALIZATION

DML dynamically adjusts the penalization based on the size of error with respect to threshold that is usually set to a percentage of the standard deviation of the actual values in the training dataset. This helps to reduce the impact of errors that are within the acceptable limit that is defined by the threshold while at the same time increasing the impact of errors that are beyond the threshold.

ii) MITIGATING LARGE OUTLIERS

The addition of this MP component helps to minimize the contribution of large outliers to the model and therefore decreases the model's error sensitivity. This helps the model to give more emphasis on the correct predictions of most of the values rather extending the curve to the extremes.

iii) BALANCING BETWEEN MSE AND MP

The values of hyper parameters α and γ are important in controlling the trade-off between MSE and MP components. This makes DML flexible in the sense that it can accommodate the nature of the dataset hence being robust in the sense that it can accommodate different distributions of errors. If α is higher, the model will try to minimize the MSE, while a higher γ will focus on the control of outliers by MP.

We formally defined DML as:

$$DML = \alpha \times MSE\left(Y, \hat{Y}\right) + \gamma \times MP\left(Y, \hat{Y}\right), \quad (3)$$

where,

- α and γ are hyper parameters which control the weights of the MSE and the MP.
- The MP is an additional penalty to find difference between predicted and actual values.

MP is formally described as:

$$MP\left(Y, \hat{Y}\right) = \frac{1}{N} \sum_{i=1}^{N} (max\left(threshold \times \left(Y - \hat{Y}\right), 0\right) + max(\frac{1}{threshold} \times \left(-Y_i, \hat{Y}_i\right), 0)\Big), \quad (4)$$

where,

- If the predicted value of target variable $\hat{Y}_i$ is less than actual value Y_i then penalty is measured using this term $(max\left(threshold \times \left(Y - \hat{Y}\right), 0\right))$.
- If the predicted value $\hat{Y}_i$ is higher than the actual value Y_i then the penalty is calculated using the term $(\max(\frac{1}{threshold} \times \left(-Y_i, \hat{Y}_i\right), 0))$, the penalty is reduced inversely in relation to threshold.
- The threshold is set to determine the lower limit and the upper limit for the prediction errors to be accepted.

This penalty function also guarantees that under-prediction and over-prediction are penalized in a different manner and thus the model cannot easily over-fit or under-fit the dataset. This is done by setting the threshold at 5% of the standard deviation of yield values, which only means that only extreme deviations are penalized.

c: RISK ADJUSTMENT LOSS (RAL)

This loss function is developed to improve the model stability by increasing the penalty for large prediction errors in comparison to small errors. This approach is useful for reducing the impact of outliers and extreme deviations which are normally the biggest causes of instability in the predictive models. This is achieved by adding the standard deviation of the predicted values into the loss function, which makes RAL adaptive in the sense that the penalty is adjusted according to the level of variation in the model's predictions.

In practical terms, this means that RAL does not only reduce the total error but also makes the model less sensitive to the data. This results in a better model because it is not just memorizing the data and is able to generalize on new unseen data. The involvement of RAL in the training process therefore enhances the stability and reliability of the model and therefore makes it ideal for use in areas where reliability is important.

The RAL integrated with MSE loss is applied to penalize larger errors more than smaller ones. The objective of RAL is to increase the predictive capability of model whereas simultaneously determining the risk and uncertainty involved.

$$RAL = \frac{1}{N} \sum_{i=1}^{N} (\left(\hat{Y}_i - Y_i\right)^2 + \alpha \times std(\hat{Y})) \quad (5)$$

where,

- $\hat{Y}$ denotes the predicted value.
- Y for the actual value.
- The α is used for weight adjustment for uncertainty.
- $std\left(\hat{Y}\right)$ represents the standard deviation of the predicted values.

Thus, by adjusting standard deviation of predictions, RAL ensures that the model is not only accurate but also stable, thereby minimizing the risk of producing unpredictable results that could compromise the decision-making process.

This approach helps to balance accuracy with uncertainty in predictions [58].

d: QUANTILE LOSS (QL)

The QL is especially useful in scenarios where it is important to measure variability or uncertainty of the predictions [41], QL aims at the quantiles of the target distribution which enables the model to make predictions that are related to different levels of risk or uncertainty. This is particularly helpful in predicting yields in agricultural production since it is important to know the range of possibilities that may happen.

i) CAPTURING UNCERTAINTY THROUGH PREDICTIVE INTERVALS

QL allows the model to estimate prediction intervals rather than just a central value, which helps in handling uncertainty in a structured way. By predicting multiple quantiles, the model provides an interval within future observations are likely to fall. This is useful when there is significant variability in data.

ii) REDUCING CONFIDENCE IN PREDICTIONS

By training a model to predict multiple quantile, QL provides a clearer picture of uncertainty, reducing the risk of over-confidence in predictions. Instead of committing to a single prediction that may be far off in highly variable conditions, the model generated probabilistic estimates, reflecting the range of possible outcomes.

iii) HANDLING ASYMMETRY IN DATA

In situations where the data exhibits skewness, QL is particularly effective. Unlike MSE, which assumes constant variance, QL adapts to varying error distributions, providing a more robust way to handle uncertain and non -symmetric data.

iv) QUANTILE LOSS FUNCTION

QL is applied to address the case of unequal penalties for over-estimation and under-estimation. By selecting a quantile (T), QL provides the flexibility to model different aspects of the yield distribution:

v) LOWER QUANTILES (T-0.25)

concentrates on the lower end of the distribution, ideal for risk adverse investors in an attempt to avoid the lower end yields.

vi) MEDIAN QUANTILE (T = 0.50)

Designed to predict the median, which is a more accurate measure since it gives equal weight to underestimations and overestimations.

vii) UPPER QUANTILES (T = 0.75)

Stresses greater returns, which can be useful to shareholders who are willing to accept greater risk for the potential of greater reward.

QL formally described in (6) as:

$$QL = \sum_{i=1}^{N} \left(T \times max\left(Y_i - \hat{Y}_i, 0\right) + (1-T) \times max(Y_i - \hat{Y}_i, 0)\right) \tag{6}$$

where T represents the selected quantile, Y_i is the actual value and $\hat{Y}_i$ is the predicted value.

The first term penalizes the underestimations when $Y_i > \hat{Y}_i$, while the second term penalizes overestimations when $\hat{Y}_i > Y_i$. The severity of the penalty is then scaled by the chosen quantile, which enables the model to focus on certain outcomes that reflect stakeholders 'risk appetite.

In training, QL is minimized to make the model's prediction to match the desired quantile of the target distribution. This process helps the model to not only predict the mean of the data but also to give an idea of the spread of the results. The efficiency of this approach is measured by applying QL on the test data in which the ability of the model to cover all the possible yield values is examined.

It is therefore important to integrate QL into the FNNIS model to address the uncertainty that is associated with the agricultural yield prediction to make it more effective and comprehensive for the stakeholders to use.

e: HYBRID AGRONOMIC EFFICIENCY LOSS (HAEL)

HAEL combines traditional MSE loss with agronomic efficiency and yield discrepancy components, aiming to optimize both prediction accuracy and practical utility in agricultural yield applications. This integration is particularly important in agricultural contexts, where the efficient use of resources such as nitrogen uptake (a most significant factor in intercropping case), it directly impacts yield outcomes.

The goal is to optimize both prediction accuracy and practical relevance by minimizing this loss during the training process. HAEL formally described as:

$$HAEL = \alpha \times MSE\left(Y, \hat{Y}\right) + \beta \times AE\left(Y, \hat{Y}, R\right) + \gamma \times YD\left(Y, \hat{Y}\right), \tag{7}$$

where,

- α, β and γ are hyper parameters.
- Y is actual yield and $\hat{Y}$ is predicted yield.
- R is representing resources referring the agronomic input.

The inclusion of Agronomic efficiency (AE) and Yield Discrepancy (YD) components ensures that model not only predicts yield accurately but also considers the efficient use of resources and practical implications of yield discrepancies.

We divided this formula into three following components:

I. MSE

II. AE is used to measure how efficiently agricultural resources are utilized. It typically involves domain-specific metrics that assess the effectiveness of input usage in achieving desired outcomes. The exact formulation can vary, but it generally involves ratios and differences that indicate resource use efficiency.

$$AE = \frac{1}{N}\sum\nolimits_{i=1}^{N} |\frac{AE_i - \hat{AE_i}}{AE_i + 1e - 8}|, \tag{8}$$

where,

$$\mathrm{AE} = \frac{Y_i}{R_i}, \hat{\mathrm{AE}} = \frac{\hat{\mathrm{Y}}_i}{R_i}$$

In our case, we considered resource R_i for agronomic efficiency is N_uptake (nitrogen uptake by crop).

III. YD considers the yield impact of the differences between predicted and actual values.

$$YD = \frac{1}{N}\sum\nolimits_{i=1}^{N} |\frac{Y_i - \hat{\mathrm{Y}}_i}{Y_i + 1e - 8}| \tag{9}$$

HAEL is a loss function that integrated multiple aspects of prediction accuracy and practical utility in agricultural practices. It is used to guide the training process by providing a comprehensive measure of model performance that the optimizer seeks to minimize.

3) CONVERSION TO GRADIENTS AND WEIGHT ADJUSTMENT

The optimizer updates the weights W and biases b by calculating gradients of the loss function with respect to the model parameters.

For a given loss function L gradients are computed as:

$$\Delta WL = \frac{\partial L}{\partial W}, \Delta bL = \frac{\partial L}{\partial b}. \tag{10}$$

Weights and Bias updated as:

$$W \leftarrow W - ɲ * \Delta WL, b \leftarrow b - ɲ \times \Delta bL \tag{11}$$

In (10) and (11), ΔWL and ΔbL are gradients of loss functions according to given parameters [59]. ɲ is the learning rate.

4) PERFORMANCE EVALUATION METRICS

For model evaluation, it is important to use a variety of performance metrics that provide a comprehensive insight of the model's efficacy. To determine which loss function is better for proposed model of yield prediction, we performed experiment with different loss functions and evaluate their performance based on relevant metrics such as MSE, MAE, R^2 for yield prediction [52], [60].

TABLE 2. The network architecture of proposed model.

Layer	Type	Input Shape	Output Shape
1	Linear	(128,44)	(128)
2	Linear	(64,128)	(64)
3	Linear	(32,64)	(32)
4	Linear	(1,32)	(1)

C. PROCESS DESCRIPTION OF FNNIS

In our neural network training process, the backward pass involves several critical steps across four layers, given in Figure 2. In each layer different size of weights and biases are used. Here we have described the operations performing in each layer:

Layer 1: In the first layer, gradients for the weights (layer1.weight, shape 128 × 44) and biases (layer1.bias, shape 128) are initially accumulated. This involves summing the gradients computed for each mini-batch during training. These accumulated gradients then undergo a backward pass using transposition (TBackward), updating the respective parameters to minimize loss.

Layer 2: The second layer follows a similar pattern, where gradients for (layer2.weight, shape 64 × 128) and (layer2.bias, shape 64) are accumulated. The accumulated gradients are propagated backward through the transposed weights. Additionally, matrix multiplication backward operations (AddmmBackward) handle the gradients associated with matrix multiplications, and ReLU backward operations (ReluBackward) adjust the gradients based on the ReLU activation function applied during the forward pass.

Layer 3: In the third layer, gradients for (layer3.weight, shape 32 × 64) and (layer3.bias, shape 32) are added. The transposed backward pass then propagates these gradients through the weights. This is followed by matrix multiplication operations to manage gradients from matrix multiplications. Finally, the ReLU backward operations adjust the gradients according to ReLU function used during the forward pass.

Fourth Layer: In the last layer, gradients for (layer4.weight, shape 1 × 32) and (layer4.bias, shape 1) are accumulated. These gradients go through a transposed backward pass and matrix multiplication operations, followed by adjustments through ReLU backward operations. The resulting cumulative gradients are then used to update the model parameters, and give the final gradients of the loss with respect to the input. This comprehensive process makes the network learn effectively by updating the gradients correctly and efficiently in the process of minimizing the loss function. Network architecture related information is given in Tables 2 and 3.

IV. EXPERIMENTAL RESULTS

The proposed research aims to evaluate and enhance the prediction accuracy of yield in pea-cucumber intercropping system using highly efficient and state-of-the-art deep learning techniques. The dataset initially contained raw data with numerous missing entries, which were filled using machine

TABLE 3. Operation types and their description.

Operation	Description
AccumulateGrad	Accumulates gradients for weights and biases during back propagation.
TBckward	Performs the backward pass for the transposed weights and biases.
AddmmBackward	Handles matrix multiplication gradients during back propagation.
ReluBackward	Adjusts gradients based on the ReLU activation function used in the forward pass.

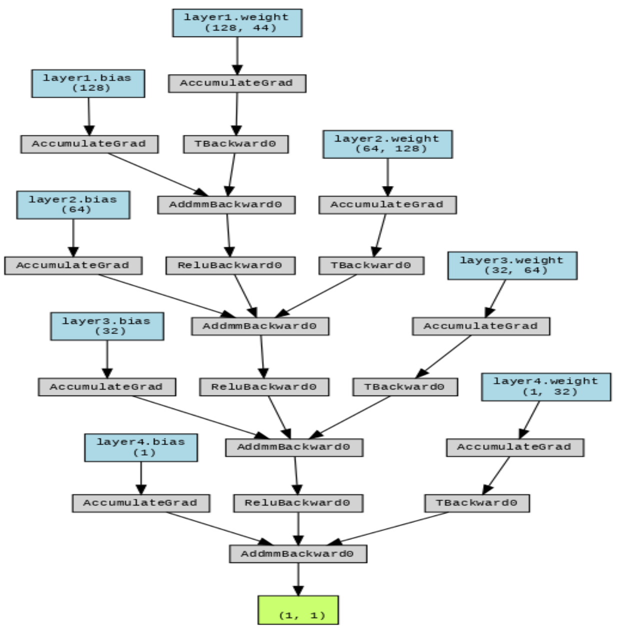

FIGURE 2. An architecture of FNNIS.

learning algorithm KNN imputation technique. All implementations were performed in Python by importing libraries such as PyTorch, Pandas, NumPy, Scikit-learn and Matplotlib for model implementation, analysis, and visualization. We used expert opinion to identify important features for yield prediction. As the dataset was based on parameters for yield prediction, we retrieved important data to simplify the dataset and removed redundant and irrelevant features.

Figure 3 presents a more detailed comparison of yield in tons/ha of pea and cucumber under intercrop and sole crop scenarios. Each dot corresponds to a yield measurement taken and cucumber is shown in green while pea is shown in orange. Trend lines for each crop represent the yield of the cropping system for the given crop. The figure also shows that cucumber has always yielded better in the intercrop system with yield varying between 1.2 to 25.7 tons/ha, while sole cropping has yield range of 1.1 to 17.9 tons/ha. The pea yields are also lower but show a better performance in the intercrop system than cucumber ranging from 0.2 to 5.2t/ha in intercrop and 0.1 to 4.9 tons/ha in sole crop.

The yields under the intercrop system are distributed more widely than those of the sole crops, which might be attributed to the complementary use of resources by the two crops. This variability is in contrast with the sole crop system where the yield distributions are less spread out, which may suggest that the resources are less flexible. These results give clear support to the hypothesis that intercropping enhances the productivity of the total system by optimizing the utilization of resources in the crops. Such findings are relevant to sustainable agriculture in the context of efficient use of land and production of crops.

A. PRINCIPLE COMPONENT ANALYSIS (PCA)

The PCA was performed to understand variability within dataset and to identify the most influential features effecting target variable yield, given in Figure 4. The analysis focused on two principal components, PC1 and PC2, which together explain 31% of the total variance. Specifically, PC1 accounts for 18.42% of the variance, which is strongly influenced by features like residue_percentN, soil_percentC, and inorganicN_mgkg. These features are indicative of soil nutrient availability and overall soil health. These findings highlight the important role of nutrient management in driving crop yield, especially in intercropping systems where competition for soil nutrients is intensified.

The results highlight that the concentration of nitrogen in crop residues (residue_percentN) and the organic carbon content in the soil (soil_percentC) are important indicators for maximizing yield. This suggests that future strategies should prioritize balancing these nutrients through appropriate fertilization and residue management practices.

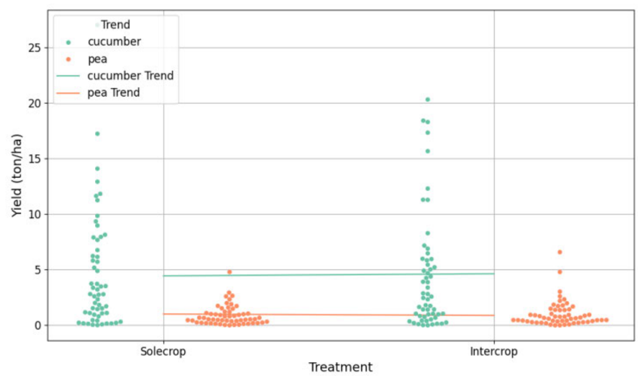

FIGURE 3. Yield (tons/ha) comparison for peas and cucumbers in intercrop and sole crop scenarios.

PC2 explains 12.65% of variance, highlights Nyield_kgha, PMN, and Fe as significant contributors to nitrogen uptake efficiency and soil fertility. These features explain how well plants can absorb available nutrients, which may affect yield.

The importance of nitrogen and micronutrients like Fe suggests that sustainable fertilization strategies should be developed to optimize these factors. This finding aligns with existing studies that emphasize the critical role of micronutrients in improving yield in poor nutrient soils, especially in intercropping systems.

The negative correlations observed between crop_percentC and N_P_ratio on PC1 Suggests that an imbalance in nutrient uptake could detract from overall yield. Similarly soil_CN and field_nut_mgmt negatively influenced PC2, indicating that improper nutrient management could limit yield potential.

This information reinforces the need for precise nutrient management interventions, particularly in managing the carbon-to-nitrogen ratio. It also implies that holistic nutrient management systems are necessary to avoid the pitfalls of nutrient competition and imbalance, especially in intercropping systems.

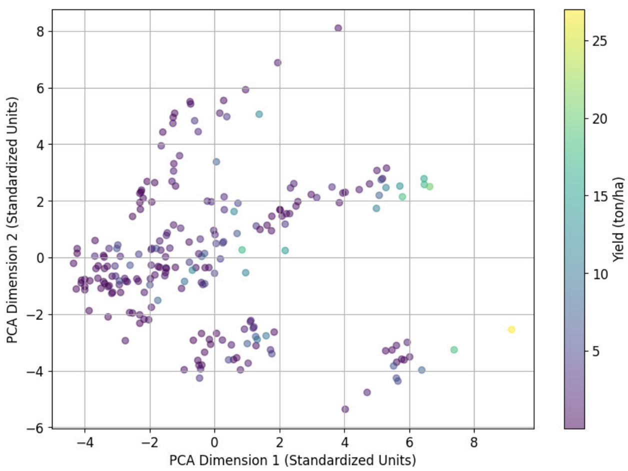

FIGURE 4. PCA scatter plot of the first two principal components. PC1 (18.42% variance) and PC2 (12.65% variance).

B. PERFORMANCE COMPARISON OF MODEL WITH DIFFERENT OPTIMIZERS

The performance of proposed loss functions is evaluated by applying three optimizers: Adam, RMSprop, and Adadelta. The results, summarized in Tables 4, 5, and 6, highlight distinct patterns in training and testing phases.

Table 4 shows the results for ANN with Adam optimizer, the MSE loss function demonstrated a relatively high performance of MAE and MSE during model training, but its testing performance was lower. DML had moderate training performance but better testing results. RAL and QL showed strong training performance with R^2 values, but a slight decline during testing. The HAEL exhibited the best training performance with the lowest MAE and MSE and highest R^2 value.

The scatter plots highlight the performance of various loss functions, including MSE, DML, QL, RAL, and HAEL, indicating their efficacy in providing accurate predictions. Figure 5 depicted all scatter plots for yield prediction by ANN for each loss functions. Points in these plots closely align with the line of perfect prediction, suggesting strong predictive accuracy. In contrast, MSE and DML exhibit few spread in their scatter plots, implying lower predictive power compared to QL, RAL, and HAEL. These observations align with the lower R-square values associated with MSE and DML, indicating their relatively weaker predictive performance.

TABLE 4. The performance of loss functions using ANN with Adam Optimizer.

Loss Functions	Training			Testing		
	(MAE)	(MSE)	(R2)	(MAE)	(MSE)	(R2)
MSE	0.0658	0.0072	0.9896	0.09837	0.0223	0.9624
DML	0.0892	0.0122	0.9847	0.0918	0.0140	0.9767
RAL	0.0338	0.0024	0.9944	0.0788	0.0135	0.9822
QL	0.0346	0.0023	0.9954	0.0814	0.0130	0.9797
HAEL	0.0564	0.0062	0.9936	0.0895	0.0129	0.9817

TABLE 5. The performance of loss functions using ANN with RMSprop optimizer.

Loss Functions	Training			Testing		
	(MAE)	(MSE)	(R2)	(MAE)	(MSE)	(R2)
MSE	0.0415	0.0131	0.9868	0.0878	0.0194	0.9608
DML	0.1150	0.0322	0.9675	0.0858	0.0123	0.9826
RAL	0.0670	0.0080	0.9919	0.1018	0.0161	0.9730
QL	0.0690	0.0107	0.9892	0.0988	0.0172	0.9757
HAEL	0.0901	0.0126	0.9872	0.0922	0.0134	0.9811

TABLE 6. The performance of loss functions using ANN with Adadelta Optimizer.

Loss Functions	Training			Testing		
	(MAE)	(MSE)	(R2)	(MAE)	(MSE)	(R2)
MSE	0.2001	0.0559	0.9411	1.1748	1.9753	1.0386
DML	0.1509	0.0599	0.9468	0.8484	1.0040	0.3136
RAL	0.1448	0.0267	0.9762	0.9025	1.1626	0.5210
QL	0.1275	0.0244	0.9783	0.8675	1.0873	0.4225
HAEL	0.0909	0.0116	0.9896	0.8700	1.0524	0.3769

In Table 5, ANN with RMSprop also performed well. MSE loss had a good balance of training and testing metrics, but not as robust as Adam. DML showed higher training MAE but excellent testing performance, showing good generalization. RAL and QL maintained strong training results but showed signs of overfitting during testing. HAEL performed well during both training and testing, though slightly less effective than Adam in some aspects.

Table 6 reveals that the Adadelta optimizer consistently showed optimal results across all loss functions.

The MSE loss function had high values for training metrics and extremely poor on testing metrics, with negative R^2 value indicating a failure to generalize. DML, RAL and QL also showed poor performance, with negative R^2 values and high testing errors. HAEL, despite good training performance had poor testing results, showing Adadelta's limitations in achieving effective optimization.

The comparison among performance metrics across three optimizer; Adam, RMSprop, and Adadelta, reveal distinct trends. Adam and RMSprop consistently demonstrate superior performance across all loss functions; whereas Adadelta's consistently poor performance showing its limitations. The low performance of Adadelta in our study is

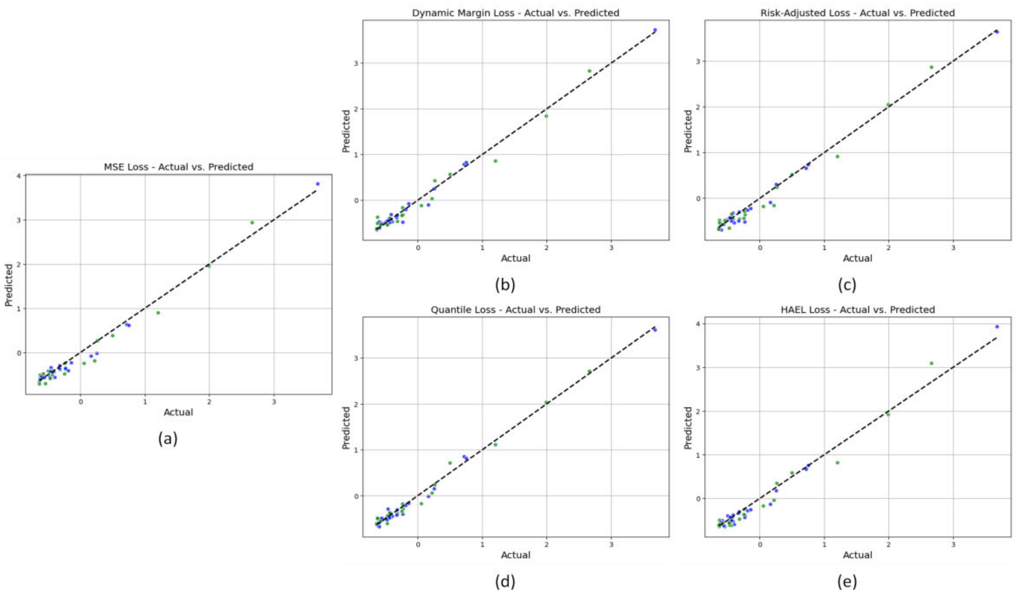

FIGURE 5. Scatter plots showing the actual vs. predicted yield (tons/ha) for the ANN model using different loss functions. (a) MSE, (b) DML, (c) RAL, (d) QL, (e) HAEL.

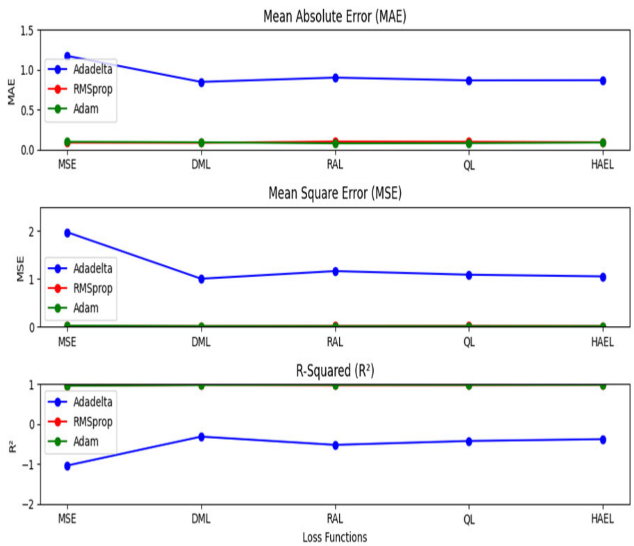

FIGURE 6. Performance comparison of three optimizers across all loss functions.

likely due to its less adaptation mechanism as compared to Adam and RMSprop. HAEL performed exceptionally well, further showing the importance of integrated loss functions in optimizing model performance.

The overall findings in Figure 6 show the comparative performance of these three optimizers. RMSprop performance was good due to low MSE while Adam optimizer show the highest R-square and shows the strong predictive power across all loss functions.

These findings emphasize the critical role of selecting appropriate loss functions and optimizers to achieve a balance between training accuracy and generalization capability, ensuring robust model performance.

C. TRAINING PROGRESS AND PERFORMANCE IMPROVEMENT OF LOSS FUNCTIONS

The training progress of the DML model is visually represented in Figure 7 for demonstrating its evolution over epochs. Initially, the model exhibits higher MP and MSE as it learns to minimize prediction errors. However, as training progresses, both MP and MSE gradually decrease, indicating the model's improvement in minimizing errors whereas adhering to defined margins. This trend shows the model's capacity to effectively balance prediction accuracy. The hyper parameters, α and γ, improves the model's yield prediction ability.

Training progress of ANN integrated with RAL over 50 epochs is given in Figure. 8. In initial epochs, the MSE exceeded 0.9, showing higher prediction errors. Uncertainty was also high at 0.575. During training, both MSE and uncertainty gradually decreased, with MSE dropped below 0.4 and uncertainty reduced to 0.45. This trend indicates that the model improves its accuracy and certainty in prediction with increasing number of epochs.

The under and over-predictions of yield by QL function during training progress are depicted in Figure 9. This figure shows a clear understanding of the distribution of prediction errors made by ANN model when integrated with the QL function. By analyzing both under predictions and over predictions, we can indicate how well the model captures the variability in yield values. The mean and median values of under predictions and over predictions shows the model's tendencies to either underestimate or overestimate yield values,

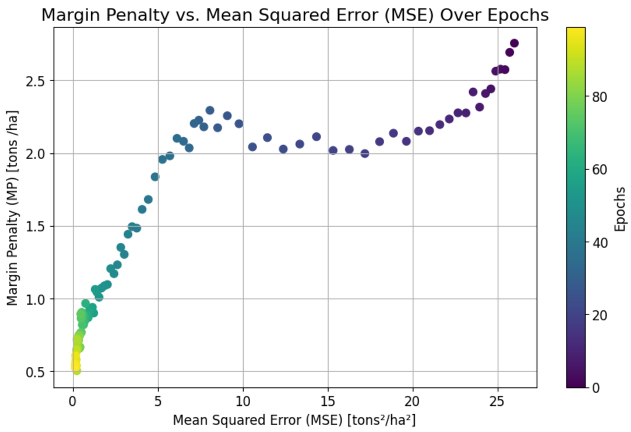

FIGURE 7. Training progress of the ANN integrated with Dynamic Margin Loss (DML).

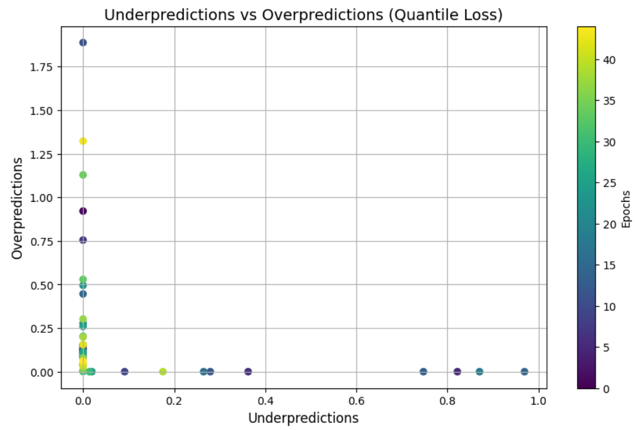

FIGURE 9. A plot showing under-prediction and over-predictions of yield (tons/ha) during model training using Quantile Loss function.

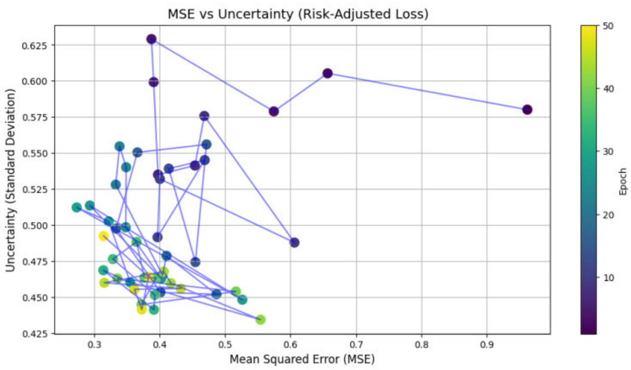

FIGURE 8. Accuracy improvement of yield prediction during training progress by using Risk Adjustment Loss (RAL).

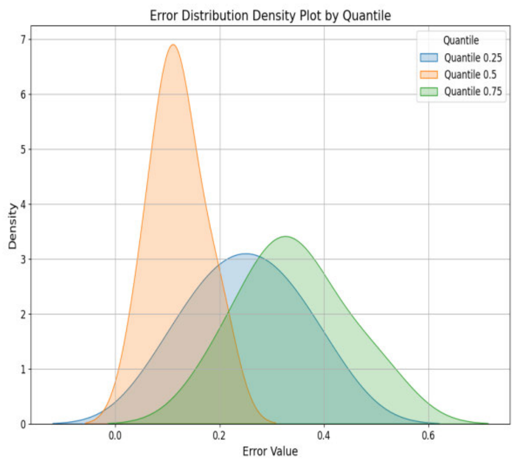

FIGURE 10. Density graph of error distribution across each quantile.

while the standard deviation captures the spread or variability of these errors, which is important to improve the accuracy and reliability of model. Overall, this analysis helps in improving the performance of model in yield predictions and improves the decision-making process for agricultural applications and yield optimization.

Figure 10 shows the density graph of error distribution across different quantiles (0.25, 0.5, and 0.75). The analysis reveals several important insights of how errors are distributed across specific quantiles, showing model's performance at various points in the data distribution.

Quantile 0.25:

- The mean error is 0.250, with a standard deviation of 0.096, reflecting a moderate spread of errors.
- The range of error values is 0.300, with a minimum of 0.100 and maximum of 0.400.
- This quantile shows a consistent error distribution but indicates some variability in predictions.

Quantile 0.5:

- The mean error drops significantly to 0.120, and the standard deviation is the lowest at 0.047.
- The error range is 0.150, with values tightly clustered around the mean.
- This suggests that the model performs more reliability at the median quantile, with minimal extreme error values.

Quantile 0.75:

- The mean error increases to 0.342, with a standard deviation of 0.093, indicating a broader distribution of error values.
- The range of error values is 0.300, with maximum error reaching 0.500.
- The wider spread and higher mean error at this quantile suggest that the model encounters more substantial prediction errors.

These findings emphasize the effectiveness of the QL function in managing error distribution across different quantiles. QL demonstrates superior performance at the median (0.5 quatile), achieving the lowest mean error and variability. However, the performance at higher quantile (0.75) requires more refinement to reduce substantial errors, highlighting the need for more model improvement. These understandings

TABLE 7. A comparison of computational efficiency and memory usage of loss functions.

Loss Function	Execution Time (s)	Memory Usage (MB)
MSE	0.0622	440.9844
DML	0.0577	440.9844
RAL	0.0588	440.9844
QL	0.0571	440.9844
HAEL	0.061	440.9844
MAE	0.0876	442.332
Huber Loss [61]	0.0668	442.332

are important for enhancing model's accuracy and reliability, supporting more accurate yield predictions.

The training progress of model using HAEL over 50 epochs, showing remarkable improvements of its key components, given in Figure 11. The MSE loss component exhibited a mean of 0.025 and a standard deviation of.015, indicating consistent and relatively low error levels throughout training. In contrast, the agronomic efficiency component shows variability. The variability highlights the model's adaptability to diverse agronomic conditions, important for accurate predictions that optimize agricultural practices. Similarly, the yield discrepancy component showing model's ability to maintain the yield prediction consistency. This proficiency in predicting yield variations is the reason that it helps in the refinement of the prediction and increase in accuracy over time. This scatter plot analysis is very useful in showing dynamic behavior of the model and emphasizes the flexibility of the model in capturing agronomic relationships. Overall, HAEL loss function is useful in minimizing MSE and improving the accuracy of agricultural yield prediction while effectively managing the agronomic efficiency and yield discrepancy control. This evaluation shows how it is useful in providing information for strategic decisions in the management and planning of agriculture.

D. COMPUTATIONAL EFFICIENCY AND MEMORY USAGE OF COMPARISON

A comparison reveals significant differences in the computational efficiency and memory usage of various loss functions (see Table 7). The comparison shows that the proposed loss functions (DML, RAL, QL, and HAEL) are highly efficient in both execution time and memory usage. DML and QL are the fastest, while all proposed functions consume the same memory as MSE. Traditional loss functions like MAE and Huber are slower and use more memory, indicating that our proposed loss functions offer better computational efficiency. This makes them suitable for large-scale agricultural yield prediction tasks, providing both accuracy and performance advantages over traditional methods.

E. LOSS FUNCTIONS SENSITIVITY

The sensitivity analysis across various loss functions reveals distinct patterns in feature importance (see Figure 12).

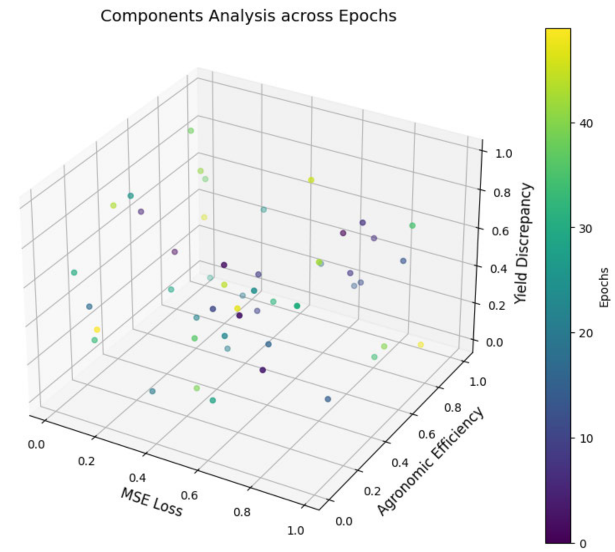

FIGURE 11. A 3D graph to illustrate the performance of each component in HAEL change over 50 training epochs.

For the MSE loss function, features such as *crop_percentN*, and *N_uptake* exhibit higher sensitivities, highlighting their significant impact on model performance. In contrast, features like *residue_N_kgha* and *crop_biomass* show lower sensitivities, suggesting they are less influential under MSE. For the DML loss function, higher sensitivities are observed for *residue_biomass_kgha*, *residue_percentN*, *Nyield_kgha*, indicating in optimization. On the other hand, features such as *crop_biomass* and *N_P_ratio* display lower sensitivities. The RAL loss function reveals higher sensitivities for *crop_percentN*, *residue_percentN*, and *soil_organic_carbon*, while features like *crop_CN* and *crop_biomass* exhibit lower sensitivities. In the case of QL, the highest sensitivities are found in *Nyield_kgha* and *nutrient_balance*, with relatively lower sensitivities in *N_uptake* and *residue_percentC*. Finally, the HAEL loss function shows notable sensitivities for *N_uptake* and *residue_biomass_kgha*, emphasizing their importance, whereas *total_biomass* and *crop_biomass* have minimal sensitivity.

F. WEIGHT ANALYSIS

The heatmap for weight analysis across different loss functions reveals varying intensities of feature importance, illustrating how each loss function emphasizes features differently in the model's performance (see Figure 13).

For the MSE loss function, *crop_percentN* and *N_uptake* were identified as the most influential features, with the highest weights. This indicates their significant role in minimizing prediction errors. Conversely, *residue_biomass_kgha* and *crop_biomass* received lower weights, suggesting they contribute less to MSE based predictions.

With DML, features such as *residue_biomass_kgha* and *residue_percentN* were weighted heavily, indicating their importance on optimizing dynamic margins. Features like

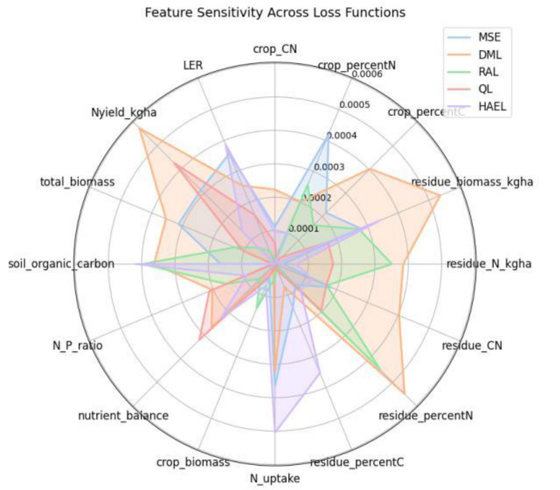

FIGURE 12. Sensitivity analysis across all loss functions.

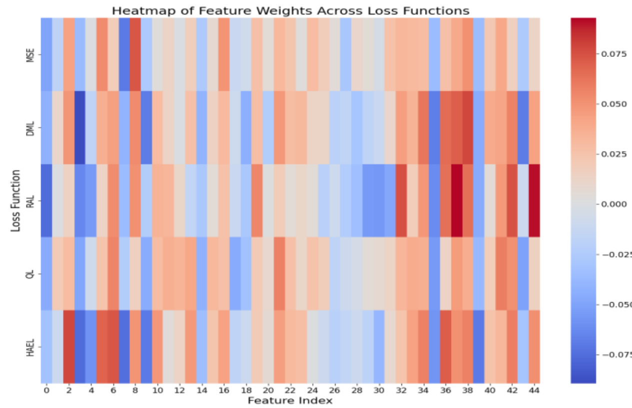

FIGURE 13. Weight analysis across all loss functions MSE, DML, RAL, QL, and HAEL.

crop_biomass and *N_P_ratio* had lower weights, reflecting reduced impact in the context of DML.

The RAL function emphasized *crop_percentN*, *residue_percentN* and *soil_organic_carbon* with higher weights, underscoring their relevance in risk adjusted predictions. In contrast, *crop_CN* and *crop_biomass* were assigned lower weights, indicating their diminished role in this framework.

For QL, *Nyield_kgha* and *nutrient_balance* emerged as key features with higher weights, reflecting their importance in modeling different quantiles of the prediction distribution. Features such as *N_uptake* and *residue_percentC* had lower weights, showing lesser significance.

The HAEL function highlighted *N_uptake* and *residue_biomass_kgha* as the most important features, emphasizing their role in optimizing both agronomic efficiency and minimizing yield discrepancies within the yield prediction model. In contrast, *total_biomass* and *crop_biomass* were assigned minimal weights, suggesting a lower contribution within the HAEL framework.

Overall, the weight analysis provides information into how different loss functions prioritize various features, helping in model refinement and enhancing the interpretability of features contributions across different loss functions.

G. ROBUSTNESS TEST

The Monte Carlo (MC) robustness test evaluated MSE with proposed loss functions, involves running 100 trials to assess the stability and reliability of a model's performance under varying conditions (see Figure 14). MC focuses on stability and effectiveness by analyzing the variance and trends in average loss.

The results are summarized as follows:

- MSE *c*onsistently showing the highest average loss across trials, indicating that it may not be effective for the model in this context. This significant variability in MSE suggests instability, which could impact model performance and generalization.
- DML also started with lower average loss compared to MSE, indicating improved performance. It's decreasing trend showing better stability and effectiveness as the trial progress.
- RAL exhibited a lower initial average loss than QL and was comparable to HAEL. It remains maintained consistent performance throughout the trials, indicating robust effectiveness in minimizing losses. RAL perform slightly better than QL, making it a strong option for reliable model.
- QL displayed effective performance with average loss lower than MSE but slightly higher than RAL. Its performance remained consistent and close to RAL, suggesting that it is a viable option with comparable effectiveness.
- HAEL started with the lowest average loss similar to RAL, indicating its strong performance in reducing losses. Like RAL, HAEL demonstrated stability throughout the trials, making it a competitive choice for minimizing loss.

These results indicate that RAL and HAEL are recommended for their robustness and performance, with QL as strong alternatives depending on specific needs.

H. CROSS VALIDATION ANALYSIS

The analysis of cross-validation results offers value able insights into how different loss functions perform across various folds.

Figures 15 illustrate the performance using MSE, MAE and R^2. Figure 15(a), shows that loss functions such as RAL and QL generally achieve lower MAE scores, indicating superior performance in minimizing error. Figure 15(b) highlights that these same loss functions consistently exhibit lower MSE values, further underscoring their accuracy in predictions. Additionally, Figure 15(c) demonstrates that RAL and QL models explain a higher proportion of variance in the data, reflecting a better model fit.

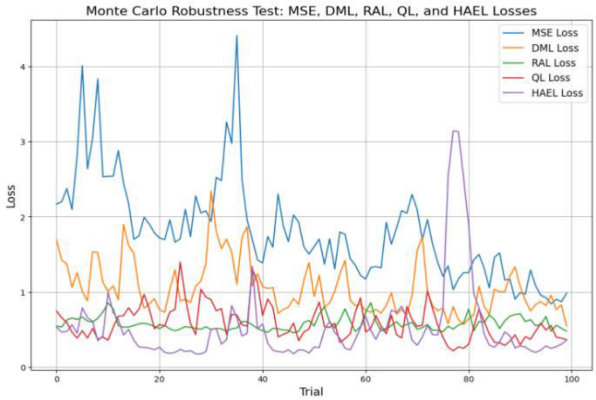

FIGURE 14. Robustness profiling of loss functions in predictive analytics.

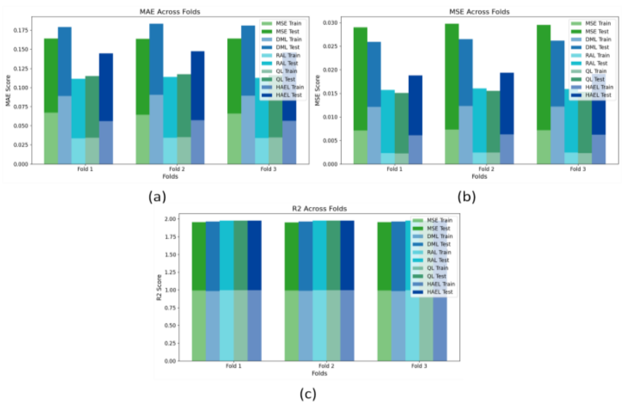

FIGURE 15. Performance of different loss functions across cross-validation folds. (a) MSE (b) MAE, and (c) R2 values for each loss functions.

Overall, the analysis suggests that RAL and QL provide more robust performance across folds, with low error metrics and high explanatory power compared to other loss functions. This comprehensive evaluation is important for assessing generalizability of these models, ensuring they perform consistently well on unseen data and are not overfitting. By confirming that these loss functions maintain robust performance across various data subsets. We enhance the reliability and accuracy of predictions in real world applications.

I. MODEL PERFORMANCE COMPARISON AND ANALYSIS WITH VARYING ARCHITECTURE MODELS

The comparison of different sate of the art machine learning models in terms of various loss functions shows the efficiency of proposed model for the given task (see Table 9).

The comparison for yield prediction shows that ANN is the best model, with QL and HAEL loss functions showing the best results in both training and testing phases, exhibiting with lowest MSE and the highest R^2 (see Table 4). In Table 9, the GBM which is the second best model performing well with slightly higher testing MSE than ANN. LSTM model shows competitive performance especially with RAL but

TABLE 8. A Performance comparison of ANN with state-of-the-art machine learning algorithms.

Model	Loss Function	Training MSE	Training MAE	Training R²	Testing MSE	Testing MAE	Testing R²
GBM	MSE	0.011	0.0723	0.9865	0.0254	0.1037	0.9553
	DML	0.0164	0.0952	0.981	0.0279	0.1114	0.9492
	RAL	0.0137	0.0687	0.9843	0.0267	0.1075	0.9524
	QL	0.0143	0.0732	0.9832	0.0273	0.109	0.9509
	HAEL	0.0151	0.0801	0.9818	0.0281	0.1123	0.9481
LSTM	MSE	0.0147	0.0932	0.9817	0.0272	0.1142	0.9468
	DML	0.0178	0.1025	0.9795	0.0301	0.1217	0.9405
	RAL	0.011	0.0789	0.9876	0.0243	0.1082	0.9531
	QL	0.0123	0.083	0.9863	0.0256	0.1115	0.9501
	HAEL	0.0182	0.1092	0.9786	0.0314	0.1239	0.9382
RF	MSE	0.06	0.0762	0.9966	0.4165	0.2368	0.970
	DML	0.0712	0.0834	0.9951	0.4471	0.2483	0.9624
	RAL	0.0581	0.0751	0.9968	0.413	0.2339	0.9707
	QL	0.0623	0.0782	0.9962	0.4278	0.2405	0.9669
	HAEL L	0.0549	0.0731	0.9973	0.404	0.2294	0.9736
CNN	SE	0.2954	0.3327	0.7241	0.3625	0.3574	0.6912
	DML	0.3104	0.3458	0.7115	0.3841	0.3682	0.6689
	RAL	0.2756	0.3242	0.7389	0.3550	0.3498	0.6951
	QL	0.2900	0.3369	0.7265	0.3718	0.3603	0.6802
	HAEL	0.2667	0.3183	0.7502	0.3495	0.3451	0.7014
SVR	MSE	6.996	0.9675	0.6093	5.647	1.0924	0.5929
	DML	7.543	1.0051	0.5982	5.9932	1.1237	0.576
	RAL	6.8701	0.9558	0.6182	5.6038	1.0815	0.6014
	QL	7.1219	0.983	0.6124	5.812	1.1085	0.5867
	HAEL	6.8054	0.9483	0.6236	5.5241	1.073	0.6079

lower than ANN and GBM in terms of testing accuracy. Random Forest (RF) provides competitive results with the HAEL function, but still lower as compared to ANN with QL, especially in terms of MSE and MAE. CNN and SVR have the lowest performance in general, and have problems with fitting and generalization. This comparison shows that ANN and GBM outperform the other models, and also reveals the weakness of CNN and SVR in this regard.

J. COMPARATIVE ANALYSIS WITH EXISTING LOSS FUNCTIONS

A comparative analysis between existing loss functions and our proposed loss functions (DML, RAL, QL, and HAEL) is provided in Table 8. Several key performance criteria are considered for analysis. Existing Loss functions like MSE are highly sensitive to outliers, whereas our proposed loss functions demonstrate improved robustness against such anomalies. In terms of gradient behavior, traditional L1 and L2 loss functions often face issues with local minima and flat regions, leading to suboptimal convergence. Our proposed loss function exhibit better convergence properties and effectively avoids flat regions in the optimization landscape. Computational efficiency is another area where proposed loss functions excel, maintaining accuracy while requiring less computational complexity compared to existing non-convex loss functions. Additionally, our proposed loss functions enhance robustness to noise, overcoming the reduced neuron activation observed with L1 and L2 loss functions.

TABLE 9. A comparison of existing loss functions with the proposed loss functions based on key performance criteria.

Criterion	Existing Loss Function	Proposed Loss Functions (DML, RAL,QL, and HAEL)
Outlier sensitivity	MSE (high sensitivity),MAE (Less sensitive) [5]	Improved robustness against outlier
Gradient Behavior	L1,L2 (possible local minima) [62]	Better convergence, avoidance of flat regions
Computational Efficiency	Complex non convex loss functions [63]	Efficient computation with maintained accuracy
Robustness to noise	L1, L2 (Reduced neurons activation) [62]	Improved noise robustness
Generalization Performance	Convex losses (Bounded performance) [64]	Enhanced generalization and consistent performance

TABLE 10. Nomenclature.

Features	Full Name	Description
crop_percentN	Crop Nitrogen Percentage	The percentage of nitrogen present in the crop.
N_uptake	Nitrogen Uptake	The amount of nitrogen absorbed by the crop.
residue_N_kgha	Residue Nitrogen (kg/ha)	The amount of nitrogen in the crop residue measured in kilograms per hectare.
crop_biomass	Crop Biomass	The total mass of the crop produced, including all parts (stems, leaves, etc.).
residue_biomass _kgha	Residue Biomass (kg/ha)	The amount of crop residue biomass produced, measured in kilograms per hectare.
residue_percentN	Residue Nitrogen Percentage	The percentage of nitrogen present in the crop residue.
Nyield_kgha	Nitrogen Yield (kg/ha)	The yield of nitrogen from crops, measured in kilograms per hectare.
N_P_ratio	Nitrogen to Phosphorus Ratio	The ratio of nitrogen to phosphorus in the crop.
crop_CN	Crop Carbon to Nitrogen Ratio	The carbon-to-nitrogen ratio in the crop.

Lastly, while convex loss functions provide bounded performance, they may not generalize well across diverse datasets. Our proposed loss functions are designed to improve generalization and maintain consistent performance. While our evaluation primarily focused on specific scenarios with three optimizers and machine learning models of varying architectures, further testing across diverse dataset would help validate their broader applicability and robustness in real-world applications.

K. NOMENCLATURE

Following terms are explained here for ready reference.

V. DISCUSSION

This research aimed at evaluating the impact of different loss functions in improving yield prediction performance of pea-cucumber intercropping systems using deep learning algorithms. The result presented in this paper provides a useful reference for the effectiveness of various strategies and can be used for further development of the agricultural modeling and optimization.

In analyzing the dataset variability, PCA revealed nitrogen and micronutrients, such as iron, as key contributors to yield enhancement. Negative correlations pointed to the fact that the balance of nutrients should be controlled very carefully to prevent the occurrence of deficiencies and to ensure that yield potential is achieved. These results are in line with the existing literature, thus supporting the need to control nutrient availability in intercropping.

The analysis of loss functions showed different performance characteristics. Although the MSE function was quite useful in the beginning, its performance was not impressive in the testing phase as compared to other loss functions. On the other hand, the DML function showed reasonable performance during training and superior performance during testing, which proves that it can be applied to different yield prediction situations. The RAL and QL functions had high training accuracy and QL achieved high accuracy in different quantiles. The proposed HAEL function, combining agronomic factors with MSE, showed the best results in both training and testing phases, which proved the effectiveness of the developed approach in predicting yield.

In the case of optimizers, Adam, and RMSprop emerged as the best performing optimizer for all the loss functions. The high values of R2 and low error rates were achieved due to their flexibility and stability for yield prediction tasks. However, Adadelta provided suboptimal performance in testing phase, which highlighted the algorithm's inability to learn complex prediction tasks. This goes to show that it is important to choose optimizers with good adaptation strategies like the Adam or RMSprop to make accurate predictions.

The training of models using different loss functions showed that the model's training progress was improving over epochs. The DML function with MSE shows progressive reduction of mean prediction errors and effective error minimization. The results also indicated that RAL significantly enhanced the performance of yield prediction by the improvement of MSE and the decrease of uncertainty. Under and over predictions were identified by QL and gave information about error distribution, while QL was strong at the median quantile, there were suggestions for improvement. HAEL's training progress reflected its capability to manage MSE, agronomic efficiency, and yield discrepancy at the same time, which proved its versatility in yield prediction.

During computational analysis, DML and QL were the fastest and most precise. On the other hand, the conventional loss functions such as MAE and Huber were slower and required more memory. The efficiency of proposed functions makes them useful for large scale yield prediction tasks where computational resources are scarce.

The sensitivity and weight analysis indicated that different loss functions focus on different features, providing a better understanding of yield prediction. For example, MSE is sensitive to nitrogen related features, while HAEL assigns greater weight to N_uptake and residue_biomass, making it more appropriate for yield prediction in diverse contexts. Similarly DML is sensitive to residual related features and RAL emphasize to nutrient balance components. Finally, the analysis enable one to understand how various characteristics affect model performance under various loss functions and stress the role of choosing the right loss functions depending on the context and purpose of the yield prediction model.

The MC simulation underscores the robustness of RAL, HAEL and QL, which provided stable and lower-loss performance from the outset, whereas MSE and DML required a learning phase to achieve similar levels of performance levels.Cross- validation results, further highlighted the stability of RAL and QL, with demonstrated superior generalization and fitting.

A comparative analysis of machine learning models (ANN, GBM, RF, CNN, SVR) with various loss functions provides valuable insights into their performance. The ANN model outperforms other models, particularly with QL and HAEL, demonstrating superior accuracy with the lowest testing MSE and highest R^2 values, making it highly suitable for complex yield prediction problems. GBM also performed well, with slightly higher testing errors. Meanwhile, LSTM showed promising performance with the RAL function but falls short in testing accuracy compared to ANN and GBM highlighting the potential limitations in sequential models for smaller database.

Finally, while advanced loss functions like HAEL, DML, RAL and QL enhance prediction accuracy by incorporating agronomic and yield discrepancy factors, they also introduce a trade-off between accuracy and interpretability. For example, HAEL offers value able understanding into agricultural practices but may reduce model interpretability compared to simpler loss functions like MSE. This balance between performance and clarity is a common challenge in machine leaning models for agricultural predictions.

VI. CONCLUSION

This study demonstrates the effectiveness of advanced loss functions integrated within FNNIS to improve yield prediction in pea-cucumber intercropping system. The conventional approaches that depend solely on MSE loss function fail to grasp the complexity and the dynamics of intercropping systems. The proposed approach is also robust to various types of prediction errors, such as risk, uncertainty, and practical relevance, with the help of DML, RAL, QL and HAEL. The comparative analysis showed that these advanced loss functions, especially when used together; improve the model's predictive ability and stability. The use of domain specific knowledge and this approach to error handling ensures that the model is better equipped to handle the complexities of intercropping systems, leading to more reliable and practical yield predictions. Thus, our results emphasize the need to choose the right loss functions and optimizers with prediction model that would be suitable for particular agricultural environments. The proposed FNNIS framework with enhanced loss functions not only enhances the yield prediction and accuracy but also contributes to the effective decision-making process in practice of agriculture. The findings of this study can open the door for the development of more complex and context-based yield prediction models, which is important step in the development of sustainable and effective agricultural systems.

LIMITATIONS

Despite the promising results, several limitations need to be pointed out:

- Hyper parameter Tuning: the loss functions developed in this work are the advanced loss functions and they are sensitive to hyper parameters, which makes the process of optimization slow and computationally expensive.
- The proposed model has been developed and validated using a pea-cucumber intercropping dataset, which restricts its applicability to other crop combinations. This may impact the generalizability of the findings to broader intercropping systems.
- The lack of publicly available datasets for other intercropping systems has limited the external validation of the model. This restricts the model's empirical validation beyond the specific case study.
- The model was trained using data from a specific geographic region (Santa Catarina), which may affect its performance in other climatic and soil conditions.

FUTURE WORK

In future, we will focus on further refinement of our loss functions and extending the testing in different intercropping systems and environmental conditions. We recognize the need to enhance the generalizability of our model, so we plan to collaborate with researchers to gather datasets that cover diverse crop combinations, beyond just pea and cucumber from various regions and climates. This will help validate our model across different agro-ecosystems. We also aim to add more features related to soil health, weather conditions and pest management to better reflect the complexities of real-world farming. We will investigate methods for enhancing the interpretability of complex models, such as HAEL, without sacrificing accuracy. Thus by enhancing accuracy of our models, we hope to participate in the development of advanced and more sustainable agricultural systems capable of addressing the demands of the modern farming.

PRACTICAL IMPLICATIONS

The advanced loss functions DML, RAL, QL and HAEL significantly improve yield prediction accuracy, supporting better decision making in agriculture. These functions can be integrated into existing systems to optimize resource use and improve yield outcomes across various cropping systems. Their practical benefits include enhanced precision in managing crop yield uncertainty and agronomic efficiency, offering value able tools for farmers and agronomists. Future work will validate their applicability across diverse crops and conditions, confirming their relevance in real world's agricultural practices.

REFERENCES

[1] J. F. Pierre, U. Singh, E. Ruiz-Sánchez, and W. Pavan, "Development of a Cereal–Legume intercrop model for DSSAT version 4.8," *Agriculture*, vol. 13, no. 4, p. 845, Apr. 2023.

[2] M. A. Raza, "Maize/soybean strip intercropping produces higher crop yields and saves water under semi-arid conditions," *Frontiers Plant Sci.*, vol. 13, Nov. 2022, Art. no. 1006720.

[3] S. Glaze-Corcoran, "Understanding intercropping to improve agricultural resiliency and environmental sustainability," *Adv. Agron.*, vol. 162, pp. 199–256, Jan. 2020.

[4] M. Weih, E. Adam, G. Vico, and D. Rubiales, "Application of crop growth models to assist breeding for intercropping: Opportunities and challenges," *Frontiers Plant Sci.*, vol. 13, Feb. 2022, Art. no. 720486.

[5] X. Zhao, H. Liu, W. Fan, H. Liu, J. Tang, and C. Wang, "AutoLoss: Automated loss function search in recommendations," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, Aug. 2021, pp. 3959–3967.

[6] B. S. Bommidi, K. Teeparthi, and V. Kosana, "Hybrid wind speed forecasting using ICEEMDAN and transformer model with novel loss function," *Energy*, vol. 265, Feb. 2023, Art. no. 126383.

[7] E. Benami, Z. Jin, M. R. Carter, A. Ghosh, R. J. Hijmans, A. Hobbs, B. Kenduiywo, and D. B. Lobell, "Uniting remote sensing, crop modelling and economics for agricultural risk management," *Nature Rev. Earth Environ.*, vol. 2, no. 2, pp. 140–159, Jan. 2021.

[8] Z. He, P. He, Y. Zhang, G. Chen, and D. Chen, "Real-time grouping of tobacco through channel weighting and dynamic loss regulation," *Ind. Crops Products*, vol. 195, May 2023, Art. no. 116427, doi: 10.1016/j.indcrop.2023.116427.

[9] J. Hu, J. Tang, and Z. Liu, "A novel time series probabilistic prediction approach based on the monotone quantile regression neural network," *Inf. Sci.*, vol. 654, Jan. 2024, Art. no. 119844, doi: 10.1016/j.ins.2023.119844.

[10] C. Xu, C. Lu, X. Liang, J. Gao, W. Zheng, T. Wang, and S. Yan, "Multi-loss regularized deep neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 12, pp. 2273–2283, Dec. 2016, doi: 10.1109/TCSVT.2015.2477937.

[11] F. Hajamohideen, "Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function," *Brain Inform.*, vol. 10, no. 1, pp. 1–13, 2023.

[12] C. MacLaren, W. Waswa, K. T. Aliyu, L. Claessens, A. Mead, C. Schöb, B. Vanlauwe, and J. Storkey, "Predicting intercrop competition, facilitation, and productivity from simple functional traits," *Field Crops Res.*, vol. 297, Jun. 2023, Art. no. 108926, doi: 10.1016/j.fcr.2023.108926.

[13] R. Xu, H. Zhao, G. Liu, Y. Li, S. Li, Y. Zhang, N. Liu, and L. Ma, "Alfalfa and silage maize intercropping provides comparable productivity and profitability with lower environmental impacts than wheat–maize system in the north China plain," *Agricult. Syst.*, vol. 195, Jan. 2022, Art. no. 103305, doi: 10.1016/j.agsy.2021.103305.

[14] X. Xiao, Z. Cheng, H. Meng, L. Liu, H. Li, and Y. Dong, "Intercropping of green garlic (Allium sativum L.) induces nutrient concentration changes in the soil and plants in continuously cropped cucumber (Cucumis sativus L.) in a plastic tunnel," *PLoS One*, vol. 8, no. 4, Apr. 2013, Art. no. e62173, doi: 10.1371/journal.pone.0062173.

[15] Z. Fu, "Maize-legume intercropping achieves yield advantages by improving leaf functions and dry matter partition," *BMC Plant Biol.*, vol. 23, no. 1, pp. 1–18, Sep. 2023.

[16] W. Xie, "Peanut and cotton intercropping increases productivity and economic returns through regulating plant nutrient accumulation and soil microbial communities," *BMC Plant Biol.*, vol. 22, no. 1, pp. 1–14, Dec. 2022.

[17] P. K. Ghosh, "Growth, yield, competition and economics of groundnut/cereal fodder intercropping systems in the semi-arid tropics of India," *Field Crops Res.*, vol. 88, nos. 2–3, pp. 227–237, Aug. 2004, doi: 10.1016/j.fcr.2004.01.015.

[18] K. M. Agboka, H. E. Z. Tonnang, E. M. Abdel-Rahman, J. Odindi, O. Mutanga, and S. Niassy, "Data-driven artificial intelligence (AI) algorithms for modelling potential maize yield under Maize–Legume farming systems in East Africa," *Agronomy*, vol. 12, no. 12, p. 3085, Dec. 2022, doi: 10.3390/agronomy12123085.

[19] B. Maestrini, G. Mimić, P. A. J. van Oort, K. Jindo, S. Brdar, I. N. Athanasiadis, and F. K. van Evert, "Mixing process-based and data-driven approaches in yield prediction," *Eur. J. Agronomy*, vol. 139, Sep. 2022, Art. no. 126569, doi: 10.1016/j.eja.2022.126569.

[20] R. Rashid, W. Aslam, R. Aziz, and G. Aldehim, "An early and smart detection of corn plant leaf diseases using IoT and deep learning multi-models," *IEEE Access*, vol. 12, pp. 23149–23162, 2024, doi: 10.1109/ACCESS.2024.3357099.

[21] E. Elbasi, C. Zaki, A. E. Topcu, W. Abdelbaki, A. I. Zreikat, E. Cina, A. Shdefat, and L. Saker, "Crop prediction model using machine learning algorithms," *Appl. Sci.*, vol. 13, no. 16, p. 9288, Aug. 2023, doi: 10.3390/app13169288.

[22] M. Malik, W. Aslam, E. A. Nasr, Z. Aslam, and S. Kadry, "A performance comparison of classification algorithms for Rose plants," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–16, Aug. 2022, doi: 10.1155/2022/1842547.

[23] S. S. Olofintuyi, E. A. Olajubu, and D. Olanike, "An ensemble deep learning approach for predicting cocoa yield," *Heliyon*, vol. 9, no. 4, Apr. 2023, Art. no. e15245.

[24] M. S. Basir, M. Chowdhury, M. N. Islam, and M. Ashik-E-Rabbani, "Artificial neural network model in predicting yield of mechanically transplanted Rice from transplanting parameters in Bangladesh," *J. Agricult. Food Res.*, vol. 5, Sep. 2021, Art. no. 100186, doi: 10.1016/j.jafr.2021.100186.

[25] P. Das, R. Banerjee, T. Ahmad, S. Devi, and G. Verma, "Artificial neural network based apple yield prediction using morphological characters," *Horticulturae*, vol. 9, no. 4, p. 436, Mar. 2023, doi: 10.3390/horticulturae9040436.

[26] M. K. A. Kadir, M. Z. Ayob, and N. Miniappan, "Wheat yield prediction: Artificial neural network based approach," in *Proc. 4th Int. Conf. Eng. Technol. Technopreneuship (ICE2T)*, Aug. 2014, pp. 161–165, doi: 10.1109/ICE2T.2014.7006239.

[27] V. Lamba, S. Hooda, R. Ahuja, and A. Kaur, "Wheat yield prediction using feedforward neural networks," in *Proc. 9th Int. Conf. Rel., INFOCOM Technol. Optim.*, Sep. 2021, pp. 1–6, doi: 10.1109/ICRITO51393.2021.9596464.

[28] P. S. M. Gopal and R. Bhargavi, "A novel approach for efficient crop yield prediction," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104968, doi: 10.1016/j.compag.2019.104968.

[29] W. Mupangwa, L. Chipindu, I. Nyagumbo, S. Mkuhlani, and G. Sisito, "Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa," *Social Netw. Appl. Sci.*, vol. 2, no. 5, pp. 1–14, May 2020.

[30] R. Rashid, W. Aslam, A. Mehmood, D. L. R. Vargas, I. D. L. T. Diez, and I. Ashraf, "A detectability analysis of retinitis pigmetosa using novel SE-ResNet based deep learning model and color fundus images," *IEEE Access*, vol. 12, pp. 28297–28309, 2024, doi: 10.1109/ACCESS.2024.3367977.

[31] X. Huang, J. Zhan, W. Ding, and W. Pedrycz, "An error correction prediction model based on three-way decision and ensemble learning," *Int. J. Approx. Reasoning*, vol. 146, pp. 21–46, Jul. 2022.

[32] L. Sang, Y. Xu, H. Long, Q. Hu, and H. Sun, "Electricity price prediction for energy storage system arbitrage: A decision-focused approach," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 2822–2832, Jul. 2022, doi: 10.1109/TSG.2022.3166791.

[33] M. Bonavita and P. Laloyaux, "Machine learning for model error inference and correction," *J. Adv. Model. Earth Syst.*, vol. 12, no. 12, pp. 1–22, Dec. 2020, doi: 10.1029/2020ms002232.

[34] S. S. Sajid, M. Shahhosseini, I. Huber, G. Hu, and S. V. Archontoulis, "County-scale crop yield prediction by integrating crop simulation with machine learning models," *Frontiers Plant Sci.*, vol. 13, Nov. 2022, Art. no. 1000224.

[35] E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, "The effect of choosing optimizer algorithms to improve computer vision tasks: A comparative study," *Multimedia Tools Appl.*, vol. 82, no. 11, pp. 16591–16633, May 2023.

[36] U. Bhimavarapu, G. Battineni, and N. Chintalapudi, "Improved optimization algorithm in LSTM to predict crop yield," *Computers*, vol. 12, no. 1, p. 10, Jan. 2023, doi: 10.3390/computers12010010.

[37] M. Yaqub, J. Feng, M. Zia, K. Arshid, K. Jia, Z. Rehman, and A. Mehmood, "State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images," *Brain Sci.*, vol. 10, no. 7, p. 427, Jul. 2020, doi: 10.3390/brainsci10070427.

[38] S. Prasher, L. Nelson, and M. Jagdish, "Potato leaf disease prediction using RMSProp, Adam and SGD optimizers," in *Proc. Int. Conf. Advancement Comput. Comput. Technol. (InCACCT)*, May 2023, pp. 343–347, doi: 10.1109/InCACCT57535.2023.10141714.

[39] J. R. Dwaram and R. K. Madapuri, "Crop yield forecasting by long short-term memory network with Adam optimizer and Huber loss function in Andhra pradesh, India," *Concurrency Computation: Pract. Exper.*, vol. 34, no. 27, Dec. 2022, Art. no. e7310.

[40] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687, doi: 10.1109/ISBI.2019.8759329.

[41] J. Dessain, "Improving the prediction of asset returns with machine learning by using a custom loss function," *Adv. Artif. Intell. Mach. Learn.*, vol. 3, no. 4, pp. 1640–1653, 2023.

[42] Y. Jung, S. N. MacEachern, and H. Joon Kim, "Modified check loss for efficient estimation via model selection in quantile regression," *J. Appl. Statist.*, vol. 48, no. 5, pp. 866–886, Apr. 2021, doi: 10.1080/02664763.2020.1753023.

[43] D. Roy, R. Pramanik, and R. Sarkar, "Margin-aware adaptive-weighted-loss for deep learning based imbalanced data classification," *IEEE Trans. Artif. Intell.*, vol. 5, no. 2, pp. 1–10, Feb. 2023, doi: 10.1109/TAI.2023.3275133.

[44] J. Yu, C. Hu, X. Jing, G. Zhou, and S. Jing, "Deep metric learning with dynamic margin hard sampling loss," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 7901–7905, doi: 10.23919/ChiCC.2019.8866028.

[45] A. N. Elmachtoub and P. Grigas, "Smart 'predict, then optimize,'" *Manage. Sci.*, vol. 68, no. 1, pp. 9–26, Jan. 2022, doi: 10.1287/mnsc.2020.3922.

[46] R. Kaur and S. K. Ranade, "Improving accuracy of convolutional neural network-based skin lesion segmentation using group normalization and combined loss function," *Int. J. Inf. Technol.*, vol. 15, no. 5, pp. 2827–2835, Jun. 2023.

[47] J. Ren, M. Zhang, C. Yu, and Z. Liu, "Balanced MSE for imbalanced visual regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7916–7925, doi: 10.1109/CVPR52688.2022.00777.

[48] A. E. Stratton, J. J. Comin, I. Siddique, D. R. Zak, L. D. Filipini, R. Rodrigues Lucas, and J. Blesh, "Assessing cover crop and intercrop performance along a farm management gradient," *Agricult., Ecosystems Environ.*, vol. 332, Jul. 2022, Art. no. 107925, doi: 10.1016/j.agee.2022.107925.

[49] (2022). *Raw Data for 'Assessing Cover Crop and Intercrop Performance Along a Farm Management Gradient*. [Online]. Available: https://zenodo.org/records/60799

[50] D. Mudassir, W. Aslam, and M. I. Lali, "Quantitative prediction of offensiveness using text mining of Twitter data," *Sindh Univ. Res. J. Sci. Ser.*, vol. 49, pp. 87–92, Mar. 2017.

[51] M. Malik, A. Ikram, S. N. Batool, and W. Aslam, "A performance assessment of Rose plant classification using machine learning," in *Computer and Information Science*. Singapore: Springer, 2019.

[52] A. Ikram, W. Aslam, R. H. H. Aziz, F. Noor, G. A. Mallah, S. Ikram, M. S. Ahmad, A. M. Abdullah, and I. Ullah, "Crop yield maximization using an IoT-based smart decision," *J. Sensors*, vol. 2022, pp. 1–15, May 2022, doi: 10.1155/2022/2022923.

[53] M. Jahn, "Artificial neural network regression models in a panel setting: Predicting economic growth," *Econ. Model.*, vol. 91, pp. 148–154, Sep. 2020, doi: 10.1016/j.econmod.2020.06.008.

[54] S. Walczak, "Predicting crime and other uses of neural networks in police decision making," *Frontiers Psychol.*, vol. 12, Oct. 2021, Art. no. 587943.

[55] M. Rahman and Md. Asadujjaman, "Implementation of artificial neural network on regression analysis," in *Proc. 5th Annu. Syst. Model. Conf. (SMC)*, Sep. 2021, pp. 1–7, doi: 10.1109/SMC53803.2021.9569881.

[56] T. Helmuth, N. F. McPhee, E. Pantridge, and L. Spector, "Improving generalization of evolved programs through automatic simplification," in *Proc. Genetic Evol. Comput. Conf.*, vol. 1829, Jul. 2017, pp. 937–944, doi: 10.1145/3071178.3071330.

[57] D. Syed, H. Abu-Rub, A. Zainab, M. Houchati, O. Bouhali, A. Ghrayeb, and S. S. Refaat, "Investigation on optimizing cost function to penalize underestimation of load demand through deep learning modeling," in *Proc. IECON 47th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2021, pp. 1–7, doi: 10.1109/IECON48115.2021.9589229.

[58] S. Syed, A. Iqbal, W. Mehmood, Z. Syed, M. Khan, and G. Pau, "Split-second cryptocurrency forecast using prognostic deep learning algorithms: Data curation by deephaven," *IEEE Access*, vol. 11, pp. 128644–128654, 2023, doi: 10.1109/ACCESS.2023.3331652.

[59] H. Gong, Y. Zhang, and Z.-H. Zhang, "Training demand prediction models by decision error for two-stage lot-sizing problems," *IEEE Trans. Autom. Sci. Eng.*, vol. 2, no. 1, pp. 1–16, Apr. 2023, doi: 10.1109/TASE.2023.3248623.

[60] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105709, doi: 10.1016/j.compag.2020.105709.

[61] S. Ma, D. Li, T. Hu, Y. Xing, Z. Yang, and W. Nai, "Huber loss function based on variable step beetle antennae search algorithm with Gaussian direction," in *Proc. 12th Int. Conf. Intell. Human-Machine Syst. Cybern. (IHMSC)*, vol. 1, Aug. 2020, pp. 248–251, doi: 10.1109/IHMSC49165.2020.00062.

[62] X. Hu, S. Wen, and H. K. Lam, "Improvement accuracy in deep learning: An increasing neurons distance approach with the penalty term of loss function," *Inf. Sci.*, vol. 644, Oct. 2023, Art. no. 119268, doi: 10.1016/j.ins.2023.119268.

[63] K. Wang, H. Pei, J. Cao, and P. Zhong, "Robust regularized extreme learning machine for regression with non-convex loss function via DC program," *J. Franklin Inst.*, vol. 357, no. 11, pp. 7069–7091, Jul. 2020, doi: 10.1016/j.jfranklin.2020.05.027.

[64] L. Zhao, M. Mammadov, and J. Yearwood, "From convex to nonconvex: A loss function analysis for binary classification," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Dec. 2010, pp. 1281–1288, doi: 10.1109/ICDMW.2010.57.

AMNA IKRAM received the B.S. and M.S. degrees in computer science from The Islamia University of Bahawalpur, Pakistan, in 2006 and 2016, respectively, where she is currently pursuing the Ph.D. degree in computer science. She is a Senior Lecturer with the Department of Computer Science and Information Technology, Government Sadiq College Women University, Bahawalpur, Pakistan. Her research interest includes crop yield maximization using AI driven decision based methodologies.

WAQAR ASLAM received the M.Sc. degree in computer science from Quaid-i-Azam University, Islamabad, Pakistan, and the Ph.D. degree in computer science degree from Eindhoven University of Technology, The Netherlands. He is currently a Professor of computer science and IT with The Islamia University of Bahawalpur, Pakistan. His research interests include performance modeling and QoS of wireless/computer networks, performance modeling of (distributed) software architectures, radio resource allocation, the Internet of Things, fog computing, effort/time/cost estimation and task allocation in (distributed) agile setups based software development, social network data analysis, DNA/chaos-based information security, and machine/deep learning models. For the Ph.D. degree, he received the Overseas Scholarship from HEC, Pakistan.