International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6401
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 11 March 2023*

Adaboost CNN with Horse Herd Optimization Algorithm to Forecast the Rice Crop Yield

M. Chandraprabha[1,2], Rajesh Kumar Dhanaraj[1]
[1]School of Computing Science and Engineering,
Galgotias University,
[2]Department of Computer Science and Engineering,
Galgotias College of Engineering and Technology,
India
cprabha.ramesh@gmail.com, sangeraje@gmail.com

Abstract - Over three billion people use rice every day, and it occupies about 12% of the nation's arable land. Since, due to the growing population and the latest climate change projections, it is critical for governments and planners to obtain timely and accurate rice yield estimates. The proposed work develops a rice crop yield forecasting model based on soil nutrients. Soil nutrients and crop production statistics are taken as an input for the proposed method. In ensemble learning, there are three categories, they are Boosting, Bagging and Stacking. In the proposed method, Boosting technique called Adaboost with Convolutional Neural Network is used to achieve the High accuracy by converting weak classifiers to strong classifiers. Adaptive data cleaning and imputation using frequent values are used as pre-processing approaches in the projected technique. A novel technique known as Convolutional neural network with adaptive boosting (Adaboost) technique is projected and can precisely handle more imbalanced datasets. The data weights are initialized; also the initial CNN is trained utilizing original weights of data. The weights of the second CNN are then modified utilizing the first CNN. These actions will be performed sequentially for all weak classifiers. An optimization algorithm called Horse Herd (HOA) is passed down in the proposed technique to find the optimal weights of the links in the classifier. The proposed method attains 95% accuracy, 87% precision, 85% recall, 5% error, 96% specificity, 87% F1-Score, 97% NPV and 12% FNR value.Thus the designed model as predicted the crop yield prediction in the effective manner.

Keywords - Adaptive Boosting (AdaBoost), Adaptive data cleaning, Convolutional Neural Network (CNN), Crop yield forecasting, Horse Herd Algorithm (HOA), Imputation, Soil nutrients.

I. INTRODUCTION

Rice is among the major food crops cultivated worldwide, that can be utilized over the people of three billion and also delivers the higher calories above 35 and below 60 percent than remaining crop. In terms of area, consumer demand, production, and rice is India's most important food crop [1]. Firstly, India is exporting rice crop of about 175.58 million tonnes every year, second only to China. Rice cultivation occupies approximately 29.50 million ha in Indian agriculture, demonstrating the crop's importance [2]. Climate and soil conditions have a large impact on crop yields. Soils supply the majority of macro and micro nutrients to crops [3].Important nutrients deficiency or soil unavailability may decrease the quantity and quality of food produced [4]. Rice requires sixteen nutrients, including main macro nutrients called N, P and K, other minor macro nutrients such as Mg, Ca &S, and Zn, Fe, Mn, Cu, B, Mo, and Cl as micronutrients. The most important nutrients required by plants or crops are Nitrogen (N), Potassium (K) Phosphorus (P), and Sulphur (S) [5].

Soil conditions can have a big impact on rice manufacture and supply all over the globe [6].Soil samples have been identified as key factors in agricultural climate variability studies due to soils' water and nutrient storage capabilities, which allow them to survive growth of crops in some years even when circumstances are detrimental[7]. Dry spell also reduced grain crop productivity by30%, and a rise in global average temperature lowered global rice yield by 3.2 percent on average [8]. As a result, forecasting rice production under current climate change scenarios is critical if the world is to be fed [9].Crop production prediction methodologies must be used in an agricultural economy if policies on food procurement, distribution, and pricing structure are to be thoughtfully developed, planned, and implemented [10].

Predicting agricultural output on a massive scale before cultivation in a timely and efficient manner is crucial for food security and organizational planning, particularly in today's constantly evolving world and international situation [11].Farmers continue to rely on traditional farming methods, which result in lower crop yields. Crop productivity can thus be increased by employing computer vision technology [12].Farmers who are unaware of yielding techniques suffer significant financial losses [13].Various methods for developing which was before prediction model based on weather factors are being developed. Discriminant function analysis [14], hybrid time series analysis [15], decision tree

[16], K-Nearest Neighbor (KNN), KNN with cross validation, Support Vector Machine (SVM) [17] and Naive Bayes (NB) are the some of the existing techniques developed to monitoring the soil properties for forecasting crop yield. However, it has certain limitations when it comes to forecasting rice crop yield in an accurate manner. Most of the limitations are about the increase in the error rate, In order to overcome these issues, an adaptive boosting model is combined with the convolutional neural network. Boosting is an ensemble learning method that seeks to enhance prediction power by training a series of weak models. Adaptive Boosting is one of the boosting technique that is utilized to merge several weak classifiers to create a unique robust classifier. This proposed AdaBoost-CNN classifies the nutrients level present in the soil and forecast the yield of rice crop.

The contribution of the proposed work is given below:

- Soil nutrients and crop production statistics datasets are in the proposed method for forecasting rice crop yield.
- Initially, the input dataset is pre-processed in order to fill the missing values present in the given dataset.
- Adaptive data cleaning and imputation using frequent values are used as pre-processing steps in the proposed method.
- Adaptive Boosting Convolutional Neural Network (CNN) is utilized to classify the nutrients level present in the soil.
- Horse Herd Optimization is used in the proposed classifier to calculate the accurate weights of the links.

The following section of the paper includes: Section II present the literature review that is related to node localization and routing in UWSN. In Section III, proposed methodology and architecture of the proposed part is present. Result and discussion part of the proposed method is present in Section IV. At last, Section V contains conclusion and future scope part.

II. LITERATURE REVIEW

The ability to predict future crop yield allows farmers and other stakeholders to make the best decisions for their crop.Several methods are used for forecasting the rice crop yield. In that, few methods are given below:

Varma, et al. [18] have suggested machine learning approaches like Support Vector Regression and Random Forest Regression. These were tested on three datasets. Statistical indicators such as Mean Absolute Deviation (MAD) and Mean Absolute Prediction Error (MAPE) were used to compare the forecasting effectiveness of the suggested models. A comparison of both machine learning approaches and the stepwise regression model has also been performed. Even so, the famous statistical method performed similarly to the two-machine learning algorithm.

Ajithkumar, B [19] have suggested two predictive methods based on weather, principal component regression and composite weather variables, which were utilized to predict production of two different rice varieties. The goodness of fit of these models was assessed utilizing the t test. The computed value of t in both models was lower than the t-critical value. To evaluate model performance, mean absolute percentage error (MAPE) estimation was utilized.

Jeong, et al. [20] had suggested a crop prototype and a deep learning prototype for different agricultural systems in South and North Korea to detect rice productivity earlier at the pixel scale..The deep learning model used pixel size benchmark rice production as target labels to leverage crop models. Future applications of the suggested strategy, which incorporates early crop production prediction also an analysis tool for the RI, may involve the use of other crop models or cutting-edge DL techniques.

Chu, et al. [21] have suggested a phenology-based time-series resemblance technique that utilizes multiple data sources to examine the variables that affect rice patterning, By combining socioeconomic factors with a cellular automata-Markov (CA-Markov) prototype. Distance from water bodies, Slope, and total temperature all had a substantial impact on rice spatial distribution. In order to calculate rice crop yield during various phonological phases, this study computed hyper-temporal satellite-derived vegetation indices from time series Sentinel-II pictures.

Oikonomidis, et al. [22] suggested a deep learning model to evaluate how well the underlying method works against specific performance metrics. This research looked at the XGBoost machine learning (ML) algorithm, as well as CNN-DNN, CNN-XGBoost, CNN-RNN, and CNN-LSTM.For the example study, use a public soybean dataset comprising 395 characteristics, like as weather, soil conditions, and also 25345 samples. In future, a combination of XGBoost with a deep learning method such as LSTM or RNN may be able to forecast crop yields more accurately using date sequence data.

Mishra, et al. [23] have suggested a straightforward ARIMA and ARMAX model. The wise ARIMAX model outperformed the simple ARIMA in terms of model and projection error. Forecasting with meteorological factors was attempted in this study using ARIMA modelling. This forecast would be useful for policy implications and the country's food security.

Nazir, et al. [24] had suggested a phenology-based algorithm and linear regression framework. In order to calculate rice crop yield during various phonological phases, this study computed hyper-temporal satellite-derived vegetation indices from time series Sentinel-II pictures. A variation of vegetation metrics were used to forecast paddy yield. The RMSE and ME analytical measures were utilized to validate the improving outcomes. Rice yield was accurately predicted using PLSR and sequential time-stamped vegetation indications. The accuracy of the PLS algorithm was not compared in this study to that of support vector machines, artificial neural networks, other geo statistics, or yield forecasting structures.

Kandan, et al. [25] had suggested a random forest approach to solve agriculture troubles by recommending a good crop and its average yield to a farmer based on climatic and agricultural parameters such as state, season, and rain fall. Crop forecasting is difficult due to a number of factors such as region, season, rainfall, and so on. As a consequence, a system is required to provide farmers with reliable crop recommendations based on meteorological conditions, production zones, and other factors.

Kumar, et al. [26] have suggested the discriminant function analysis (DFA) to predict rice agricultural output. DFA is used extensively to find the groups differences by calculating the optimal variables. This can be achieved due to its simple computational process. Otherwise, we can say that, when the groups are formed naturally, DFA helps to determine its membership values. According to the factors of meteorological area, this research was to foretell the yield of rice crop using pre-harvest techniques. The 15 days weather data with five weather variables like range of temperature (Minimum or maximum), average relative humidity, day hours and accumulated rainfall. The current study's findings will assist policymakers and other stakeholders in making sound decisions.

Islam et al [27] projected neural network approach called an artificial neural network which can be used for selecting the crop and predicting the yield. Accuracy and error rate was compared to achieve better output and prediction by using this algorithm. In this study, the algorithms like random forest, support vector machine and logistic regression, were implemented. 46 parameters were collected, such as temperatures in maximum and minimum range, type of soil and weather, average precipitation, soil response, humidity, structure of soil, composition of soil , soil moisture level, consistency of soil, climate required, and soil texture into this prediction process.

Guo, et al. [28] had suggested an artificial neural networks and partial least squares regression for prediction of rice yield in east china based on climate and agronomic traits data. The PLSR model showed that covariates occurred among the parameters, and modifications should be considered for climate data-based modelling. The FFBN model demonstrated better prediction performance than that of PLSR. The optimum architecture of the FFBN consisted of one hidden layer with 29 neurons. Therefore, the FFBN algorithm was an effective option for the prediction of rice yield in complex systems of rice production.

Bondre, et al. [29] had suggested a machine learning algorithms for prediction of crop yield and fertilizer recommendation. Machine learning was an emerging research field in crop yield analysis. Different machine learning techniques were used and evaluated in agriculture for estimating the future year's crop production. This study suggests and implements a system to predict crop yield from previous data. In future implement smart irrigation system for farms to get higher yield.

Lizumi, et al. [30] had suggested a global crop yield forecasting using seasonal climate information from a multi-model ensemble. Here, assess the reliability of global within-season and pre-season pre-dictions of yield variability obtained by applying statistical yield models to seasonal temperature and precipitation hindcast data derived from a multi-model ensemble (MME). This analysis was performed for five individual atmosphere-ocean coupled general circulation models (GCMs) and the two MME datasets generated using the average method and the mosaic method.

Pagani, et al. [31] had suggested a high-resolution rice yield forecasting system based on the integration between the WARM model and remote sensing (RS) technologies was developed. RS was used to identify rice-cropped area and to derive spatially distributed sowing dates, as well as for the dynamic assimilation of RS-derived leaf area index (LAI) data within the crop model. In particular, RS allowed reducing the uncertainty by capturing factors not properly reproduced by the simulation model.

Kour, et al. [32] had suggested an Integrated Moving Average with Exogenous variables (ARIMAX) time-series model along with its estimation procedure. In the present investigation, two models at tilling stage of rice growth are developed by including the most important weather variables. As an illustration, ARIMAX models are employed for forecasting of rice yield in kheda district of Gujarat. Comparative study of the fitted models was carried out from the viewpoint of Relative mean absolute prediction error (RMAPE) Mean absolute deviation (MAD) and root mean square error (RMSE) values.

According to literature, numerous systems are designed for forecasting rice crop yield. Based on the above mentioned articles several significant challenges arise during crop yield forecasting. Most of the existing researches are based on the image or spatial data's where Convolutional neural networks, Artificial Neural networks, Deep Belief Networks, Deep Neural Networks and so on are used to predict the yield for the crop like wheat, sorghum or maize. These prediction models worked well but it has limitations while handling incomplete datasets. Also, existing approaches provided very less support to the nominal data values. Popular statistical approach performance was considered [18], suitable for particular area [20], suggested methods does not compared with existing techniques [24], low accuracy and high error are some of the issues in the existing algorithms. So the proposed method uses adaptive boosting CNN classifier to classify the nutrients level of the soil and forecast the rice crop yield with high accuracy along with optimization technique.

III. PROPOSED METHODOLOGY

For organizational planning and food security, large-scale agricultural production forecasting is essential, particularly given how quickly the world and the worldwide situation are changing right now. The proposed method develops an adaptive boosting neural network to forecast the yield of the rice crop by monitor the nutrients level present in the soil.

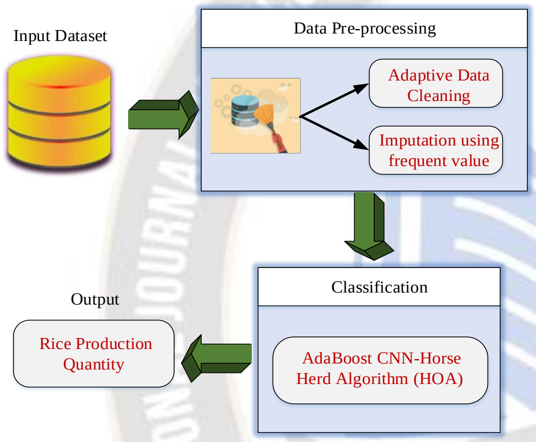


Figure 1. Architecture of Proposed Method

The proposed method work flow is illustrated in figure1. Soil nutrients [33] and crop production statistics [34] are taken as the input for proposed method. The input dataset contains large number of missing value, so it is necessary to pre-process the given dataset. In this proposed approach, adaptive data cleaning and imputation using most frequent data are used as pre-processing techniques. The pre-processing data are then fed into the classifier for classifying nutrients level present in the soil to forecast the amount of yield. Adaptive Boosting Convolutional Neural Network (AdaBoost CNN) is used as a classifier in the suggested approach. Horse Herd Optimization Algorithm (HOA) is used in the proposed approach to find the accurate weights of the links present in the classifier. This AdaBoost classifier used in the proposed approach classifies the nutrients present in the soil and forecast the amount rice crop yield in the particular soil. Figure 2 shows the flowchart of the proposed methodology.

A. *Data Pre-processing*

The given input soil nutrients and crop production statistic dataset contains some missing values. Missing data is defined as the values or data that is not stored for some variables in the dataset. Past data might get corrupted due to improper maintenance, Observations are not recorded for certain fields due to some reasons are some of the reason for the missing data. Pre-processing technique is used to fill all those missing values in the input dataset. Adaptive data cleaning and imputation using most frequent data are used in the proposed method for pre-processing the input data. Adaptive data cleaning is used to filter the noises out from the raw data because it contains noisy data. Imputation using most frequent data approach used in the proposed method to fill the missing values present in the given dataset.

i. Adaptive Data Cleaning

The original dataset comprises a lot of low-quality data with a lot of meaningless elements, while high-quality data is made by filtering out those worthless values for a specific application. We utilized Adaptive data cleaning to filter the noises out of the raw data because it contains noisy data. Adaptive data cleaning contain prediction method termed denoising auto encoder mechanism.

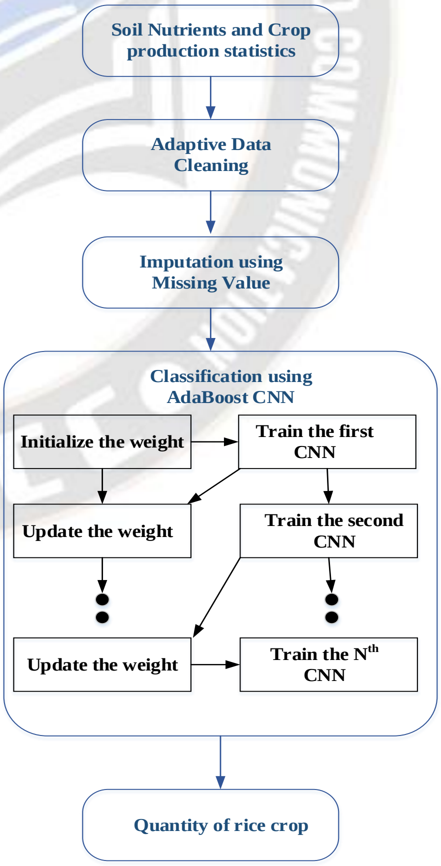


Figure 2. Flow chart of the Proposed Method

ii. Denoising Auto Encoder

Denoising Auto encoders are a basic variation of traditional auto encoder neural networks that are trained to denoise an artificially distorted version of their input rather than reconstructing it. The DAE block improves the network's overall stability.

DAE consist of input layer with data $a_t \in R^d$, input layer with noisy data $\tilde{x}_t \in R^d$, and a data in hidden layer $c_t \in R^d$ and a data in output layer $b_t \in R^d$ and it also consist of two parts that are encoder and decoder. DAE adds noise to the input data as given in eqn (1), which is the main difference between it and a regular auto encoder [35].

$$\tilde{a}_t \sim q_D(\tilde{a}_t|a_t) \qquad (1)$$

The function $m_z$can be referred as an encoder that converting a d-dimensional $\tilde{a}_t$ into a d'-dimensional z_t as shown in eqn (2).

$$z_t = m_z(\tilde{x}_t) = a_z(W_z\tilde{x}_t + b_z)$$

(2)

Where, the activation function is denoted as a_z, and the characteristics parameter of the encoder are referred as b_z,W_z. Then m_y is used as a decoder function that maps z_t back to the input space y_t that shown in eqn (3),

$$y_t = m_y(z) = a_y(W_y z_t + b_y)$$

(3)

Where the decoder's activation function is denoted as a_y and the decoder characteristics parameters are refereed as W_y and b_y.

The distance between x_t and y_t is minimized by using the loss function of DAE that is shown in eqn (4)

$$L_{DL} = \frac{1}{m}\sum_{i=1}^{m}\left\|x_t^i - y_t^i\right\|^2$$

(4)

Where L_{DL} also seeks to reduce the greatest absolute error as well. As a result, the error function has a regularize with a weight of $\eta_2 = 0.05$.

iii. Imputation Using Most Frequent Data:

The approach of substituting data with statistical estimates of the missing values is referred to as imputation. Any imputation technology's objective is to produce a complete dataset. In this proposed pre-processing approach, the missing values are replaced using frequent value present in the dataset by calculating the mean and mode. Columns in the dataset with no data can be replaced with the column's frequent values. This method can help to avoid data loss. Replacing the frequent value is a quantitative way of dealing with incomplete data.

B. AdaBoost-CNN Classifier for Rice Crop Yield Forecasting:

The pre-processing data are then fed into the classifier for classifying nutrients levels present in the soil to forecast the amount of rice crop yield. Boosting is an ensemble learning method that seeks to enhance prediction power by training a series of weak models, each of which compensates for the shortcomings of its predecessors [36].Boosting is comprised of two major algorithms: adaptive boosting (AdaBoost) and gradient boosting [37]. In this proposed method AdaBoost is used for classification purpose. AdaBoost techniques are used to merge numerous weak classifiers into a one robust classifier. In this strategy, a team of weak classifiers are trained one after the other. Each training sample is assigned a weight to reflect how much it was trained with a weak classifier, and each classifier is trained with an emphasis on the shortcomings of the prior classifier. If the previous weak classifier correctly trained a sample, its weight is reduced exponentially. Based on the results of the preceding classifier, every subsequent poor classifier is trained with additional weights for samples that weren't properly learned [38].

The multi-class AdaBoost technique is utilized in this paper to create an AdaBoost technique for CNN. This suggested new approach is termed Adaptive Boosting Convolutional Neural Network (AdaBoost CNN). (x_1, y_1)…. (x_n, y_n) denotes the training dataset, where x_n represents an input vector, y_n denotes the result corresponding to x_n ,also $y_n \in \{1, 2, …, K\}$, where K represents the total classes. To match a classifier, the training objective is to utilize the training data C(x). The trained classifier produces a data weight vector called $D = \{d_i\}$ for every sample present in training data, where i= 1, 2,..n is the number of training samples.

$d_i = {}^1\!/_n$ is used to initialize the data weights.The M networks are then successively trained. The first iteration of the sequential learning method involves random activation of the CNN weights, which are then trained for one or more epochs depending on how challenging the learning problem is. All training data are utilized to train the initial CNN, $C^{m=1}(x)$, with a weight of ${}^1\!/_n$.There are no differences in the initial CNN's relevance, i.e. the weights of various training samples. Following training, the CNN result for training samples is determined. For every input sequence, AdaBoost CNN generates a K-dimensional vector result. Estimated results for the K classes are contained in the vector. In the vector outcome, each element is a real-valued, confidence-rated forecast linked to a class. Eqn (5) depicts the vector result for an input sequence.

$$P(x_i) = [p_k(x_i)], \qquad k = 1, \dots K \qquad (5)$$

This displays the chances that one of the K classes has the forced input. The category with the greatest probability receives an input when it is tested. Initial CNN's result is shown in eqn (6).

$$P^{m=1}(x_i) = [p_k^{m=1}(x_i)] \qquad (6)$$

This is done to change the weights of the data, $D = \{d_i\}$ by the following eqn (7)

$$d_i^{m+1} = d_i^m \exp\left(-\alpha \frac{K-1}{K} Y_i^T \log\left(P^m(x_i)\right)\right), i = 1, \dots, n \qquad (7)$$

Where d_i^mrepresents the ithdata weight utilized by the mth CNN, denotes a learning rate, Yi is the ith training data label vector, and $P^m(x_i)$represents the mth CNN’s vector outcome in response to the ith training sequence. It is used to refresh the specimen weights of a CNN. If the output label $Y_i^T$and the log of the mth CNN's output sequence, $P^m(x_i)$, are correlated and their partial derivative has a large value, the exponential function in eqn (7) has a lower number. Since the current outcome is near to the label vector and indicates that the training samples have been trained by the present CNN, the smaller number of the exponential function decreases the weight of the training dataset for the following CNN. Divided by the total amount of the weights, all training data weights for the present CNN are calculated. Additionally, the next CNN's learning process begins as the trained CNN is stored.

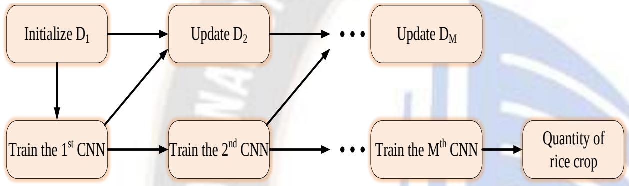


Figure 3. Workflow of the suggested AdaBoost-CNN

The proposed classifier is depicted schematically in Figure 3.The data weights are activated utilizing D_1 ={d_i = $1/n$}, and the initial CNN is trained using the first data weight.Then, using the initial CNN, C^1(x), the second CNN data weights, D_2={d_i}, are updated.Moving the trained C^1(x)to the next CNN follows.The MthCNN, C^M(x) is trained by repeating this technique (x).The classifier for that iteration is first learned utilizing training data also associated data weights, D={d_i}, in each iterative procedure of the sequences active learning. The data weights are then changed for the following iteration according to the results of the trained classifier. M weak classifiers do these two processes one after the other. After training the M classification model, testing of the AdaBoost-CNN has begun. The following eqn (8) is utilized to anticipate the classifier's result.

$$C(x) = \operatorname{argmax} \sum_{m=1}^{M} h_k^m(x) \qquad (8)$$

Horse herd optimization is used in the proposed method to find the accurate weights of the links in the classifier.

C. Horse Herd Optimization:

The Horse Herd Optimization Algorithm (HOA) is a revolutionary technique adopted by the hierarchical structure of a horse herd. The ordered arrangement of horse herds has influenced HOA. Horses can live in herds. Numerous animals that live in big crowds and it must be established in a stable hierarchy structure or "pecking order" in order to minimize violence and enhance team unity. This is a linear system most of the time, but not always. Horse B may be dominant over horse B in non-linear structures. Determine supremacy using a list of conditions that includes an individual's need for a specific resource at a specific time.As a result, it can change over the lifespan of an individual animal or a herd. Some horses may command all resources, while others may command none. It's important to remember that this isn't normal horse behavior. It is the outcome of humans putting horses in close quarters with limited resources. Horses with poor social skills are referred to as "dominant horses". Horses behave hierarchically in groups, with the more senior horses taking the first sips of water and bites of food. Because they might not get enough food, low-status animals consume last, while higher-status horses might skip meals altogether if there isn't enough food [39]. The steps of the HOA for this proposed work is as follows:

Step1: Initialization

Initialize the weight as an input,

$$\text{weight} = \{W_1, W_2, \dots . W_n\} \qquad (9)$$

Step2: Fitness Function

Error is computed here to determine the fitness value.

$$E = \frac{1}{2} \sum_{n=1}^{N} (y_n - z_n)^2 \qquad (10)$$

Where, the predicted value is represented asy_n , z_n is represented as the actual value and E is represented as the Mean Squared Error (MSE).

Step 3: Updating the value

Each iteration values are updated to find the best optimal value of the transformer. Updation of each iteration solution is done by using the below expression,

$$U_{x,y}^{T+1} = U_{x,y}^{T} + F_{x,rank} \times \left(P_{center,y}^{T} - Z_{x,y}^{t}\right) \qquad (11)$$

$$U_{x,y}^{T+1} = U_{x,y}^{T} + Rand \times \left(P_{center,y}^{T} - Z_{x,y}^{t}\right) \qquad (12)$$

Where, Rand signifies random number [0, 1]. $T+1$signifies new iteration, T represent current iteration. $Z_{x,y}^t$represent random number between 0 and 1.

Step 4: Termination

Termination is the last phase, and it happens once the best answer has been found.

This AdaBoost-CNN[40] classifier with HOA optimization classifies the nutrients present in the soil and determine how much amount of rice crop is predicted in the particular soil. Pseudo codefor the proposed method is given in Table I.

TABLE I. ADABOOST CNN WITH HORSE HERD OPTIMIZATION PSEUDOCODE

Pseudo code for proposed research (AdaBoost CNN with Horse Herd Optimization)

```
Input: Soil nutrients and crop production statistics

Step 1: Initialize the data preprocessing using Adaptive data cleaning and Imputation using eqn. (1) to eqn. (4)
Step 2: Repeat step 1 until all omitted or missed data's are preprocessed.
Step 3: Initialize the Classification algorithm called Adaboost CNN
If number of estimates m==1, then first CNN $C^{m=1}(x)$ is needed to be trained upon the training data that uses initial weights of the samples, $D_m = \{d_i = {}^1/_n\}$.
    Else
Learning parameters of the preceded CNN, $C^{m-1}(x)$ is to be transferred to the succeeding $m^{th}$ CNN called $C^m(x)$ and training to be done on the $m^{th}$ CNN for one epoch, upon the training data using the vector weight of the sample $D_m = \{d_i\}$.
Step 4: For all K number of classes, the output of the $m^{th}$ CNN called probability of class estimates, is returned $p_k^m(x)$ where k is the input probabilities that belongs to K number of classes which is represented as $k = 1,2,\dots,K$.
Step 5: According to $p_k^m(x)$, weight of the data sample $D_m$ is updated using eqn. (7).
Step 6: Updated weights of the data sample, $D_m$ is to be re-normalized.
Step 7: $C^m(x)$, $m^{th}$ CNN is to be saved for further iterations.
Step 8: Repeat Step 3 to step 7 till the completion of all M networks

#Horse Herd Optimization

Step 1: Initialize the weight using Eq (5) by calculating error values and save the best weight and its position
Step 2: while (iteration<maximum iteration)
        Select weight
        Determine the Fitness function using Eq (6)
        Update the solution using Eq (7)
        Return the best solution
end while
Output: Forecasting rice crop yield
```

IV. RESULT AND DISCUSSION

The proposed AdaBoost-CNN for forecasting rice crop yield is tested on python 3.8 with the requirements of CPU: Intel core i5 processor, GPU specification: NVidia GeForce GTX 1650, 16 bit operating system, RAM: 16GB. Soil nutrients dataset [20] and crop production statistics [21] are taken as the raw data for the proposed method. Different nutrients level of the soil is present in the soil nutrients dataset. pH, Electrical Conductivity (EC), Iron (Fe), Organic Carbon (OC), Nitrogen (N), Manganese (Mn), Boron (B), Phosphorus (P), Sulfur (S), Potassium (K), Copper (Cu), and Zinc (zn) are soil parameters. Amount of crop yield in the particular area data's are present in crop production statistics dataset. Totally 11227 number of samples were taken as the input for the proposed method.

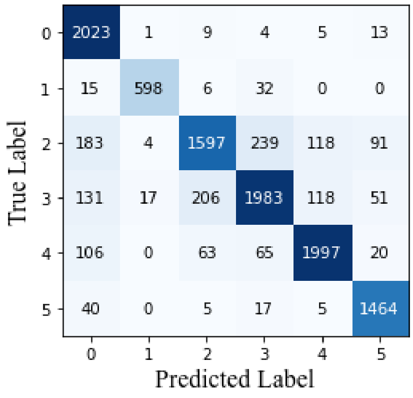


Figure 4. Confusion Matrix Plot for the Proposed Approach

Proposed method confusion matrix is illustrated in figure 4.To validate the accuracy of classification technique confusion matrix is used. Confusion matrix is a plot between the predicted label and true label. Six different classes based on crop production is consider in the proposed method. For class0, 2023 samples are accurately predicted, 598 samples are correctly predicted for class 1, for class2, 1597 samples are accurately predicted, 1983 samples are correctly predicted for class 3, for class4, 1997 samples are accurately predicted and 1464 classes are predicted correctly for class 5.

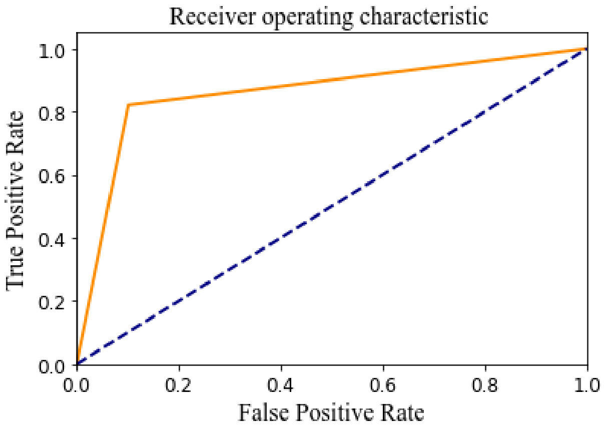


Figure 5 AUC and ROC for the Suggested Method

AUC plot and ROC plot of the suggested approach is depicted in figure 5. The Area under the Curve (AUC) and

Receiver Operating Characteristics (ROC) curve is utilized to validate or demonstrate the multi-class categorization performance. The degree or amount of separability is represented by the AUC. A great model has an AUC close to zero. In the proposed method, AUC plot is close to 1, so the proposed method has high level of severability. The ROC curve is a probability curve that demonstrates how effectively the model can discriminate across classes.

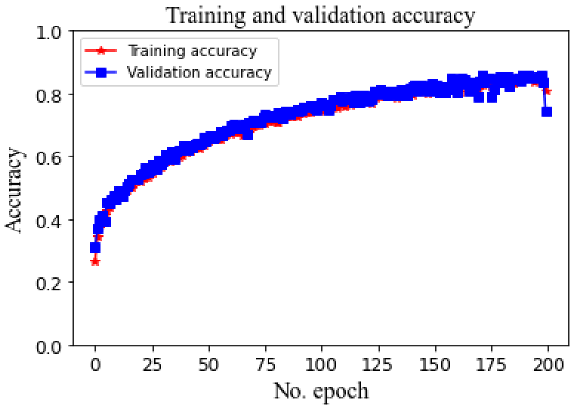


Figure 6 (a). Training and Validation in terms of Accuracy

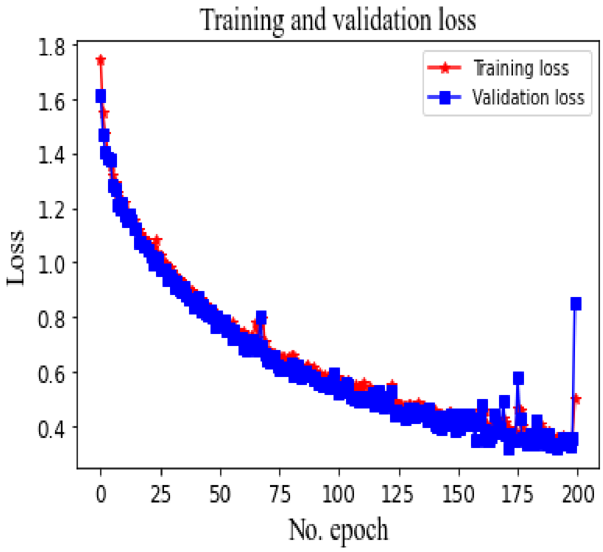


Figure 6(b). Training and validation in terms of Loss

Figure 6 (a) illustrates the training and validation accuracy of the suggested method. Training accuracy means that identical data's are used both for training and testing, while test accuracy represents that the trained model identifies independent images that were not used in training. The red and blue line in the graph indicates the training and validation accuracy respectively. When the number of epoch increases, the training as well as the validation accuracy also, is increased. Training and validation loss of the suggested approach is represented in figure 6 (b). The training loss represents how prediction model fitting with the training data, while validation loss represents how the new data can be fitted by the model. The red and blue line in this graph represents the training loss and validation loss respectively. When the number of epoch increases the loss gets decreased.

A. Comparison Analysis:

The proposed using Adaptive Boosting Convolutional Neural Network (AdaBoost CNN) classifier is compared with some existing classifiers like CNN, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN) and Deep Belief Network (DBN). The performance attained using these existing techniques are compared with the proposed AdaBoost CNN approach. Some of the performance indicators utilized for comparison between suggested and current approach are Error, Precision, Accuracy, False Negative Rate (FPR), False Predictive value (FPV), Recall and Specificity.

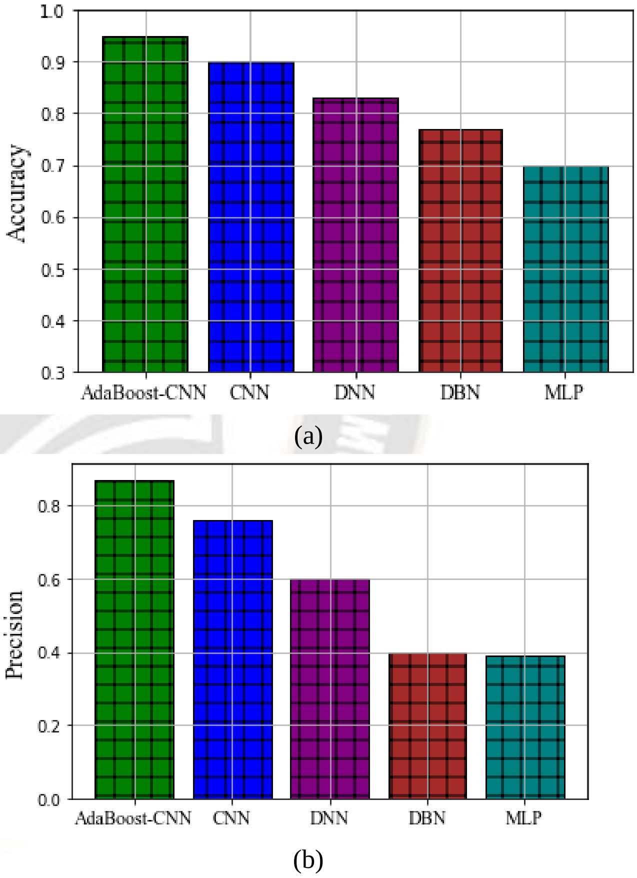


Figure 7. a) Accuracy b) Precision Metrics Comparisons

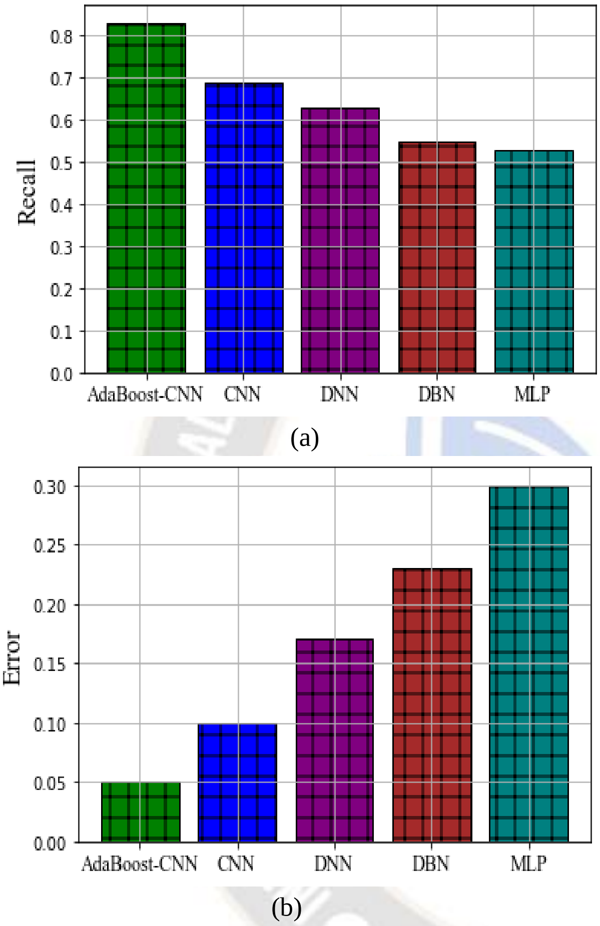
Accuracy metric comparison of the proposed and existing classifiers is illustrated in figure 7 (a). The accuracy is defined as how close it is to the genuine value. The accuracy of the suggested approach using AdaBoost-CNN classifier is 95%. But the accuracy rate achieved by using the existing classifiers like CNN, DNN, DBN and MLP are 90%, 83%, 77% and 70% respectively.

$$\text{Accuracy rate}=\frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

DOI: https://doi.org/10.17762/ijritcc.v11i4.6401
Article Received: 08 February 2023 Revised: 04 March 2023 Accepted: 11 March 2023

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (14)$$

Figure 7 (b) represents the precision metrics comparison of the proposed and existing classifiers. The precision value obtained by proposed classifier is 87%. The precision value obtained by existing CNN, DNN, DBN and MLP classifiers are 76%, 60%, 40% and 39% respectively. Accuracy and precision metrics comparison shows that the proposed approach is superior to the existing methods.



(a)

(b)

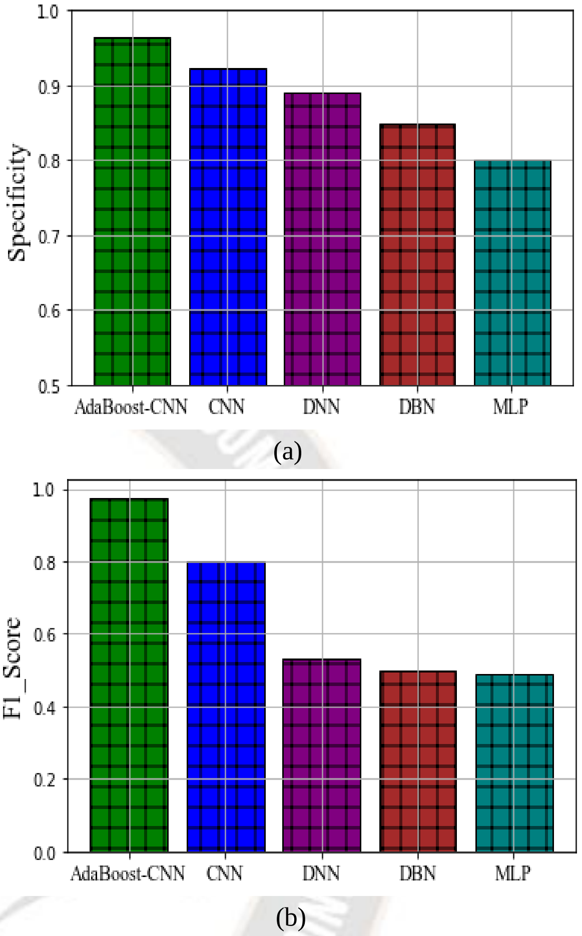
Figure 8. Comparison of a) Recall b) Error Metrics

Figure 8 (a) illustrates the recall metrics comparison of the proposed and existing classifiers. AdaBoost-CNN used in the proposed reaches 85% recall value. But the CNN, DNN, DBN and MLP existing classifiers has a recall value of 69%, 63%, 55% and 53% respectively.This demonstrates that the suggested approach has a high recall than current classifiers.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (15)$$

$$\text{Error} = 1 - \text{Accuracy rate} \qquad (16)$$

Error metrics of the comparison of the proposed and existing classifiers is shown in figure 8 (b).The error value produced by the CNN, DNN, DBN and MLP classifiers are 10%, 17%, 23% and 30% respectively. The proposed method using AdaBoost-CNN has an error value 5%. It is seen to be low when compared with existing classifiers. It shows that the proposed classifier is superior to the existing classifiers.



(a)

(b)

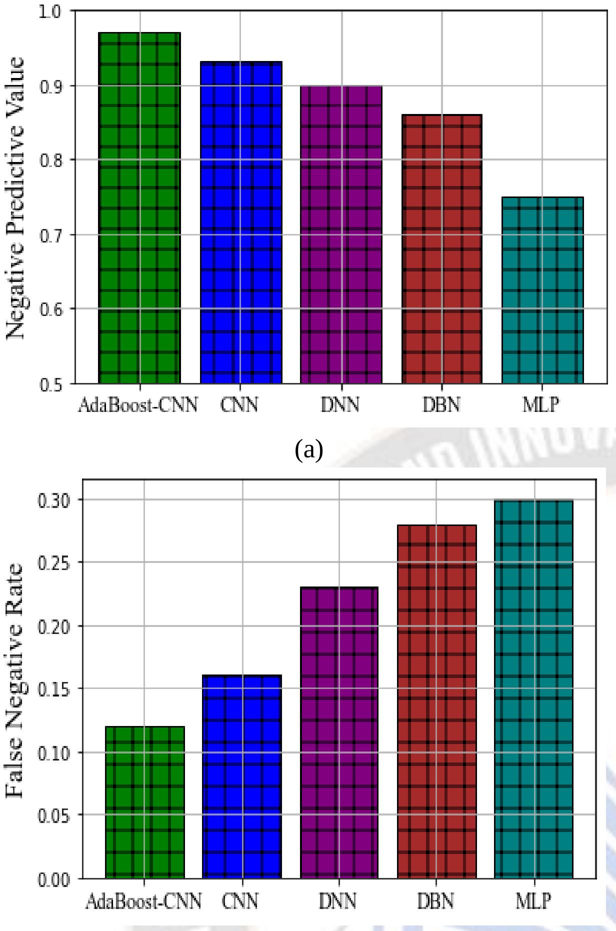
Figure 9. a) Specificity b) F1-score Metrics Comparisons

Comparison of proposed and existing classifiers in terms of specificity metrics is shown in figure 9 (a). The value of specificity for the proposed method using AdaBoost-CNN classifier is 96%. In the existing methods, the specificity value obtained by using CNN, DNN, DBN and MLP classifiers are 92%, 89%, 85% and 80% respectively.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \qquad (17)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (18)$$

Figure 9 (b) illustrates the F1_Score value of the suggested and existing classifier. F1_Score of the suggested approach using AdaBoost-CNN classifier is 87%. But the existing technique has an F1_Score value of 80%, 53%, 50%, and 49% for CNN, DNN, DBN and MLP respectively. It is shown that the proposed classifier is superior to existing classifiers.



(a)

(b)

Figure 10. Comparison of a) NPV b) FNR Metrics

Negative Predictive Value (NPV) metrics comparison of the proposed and existing classifiers is shown in figure 10 (a).AdaBoost-CNN used in the proposed method reaches 97% NPV. But the CNN, DNN, DBN and MLP existing classifier has an NPV of 93%, 90%, 86% and 75% respectively. This shows that the NPV of the suggested approach is higher than the current approach.

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN+FN}} \qquad (19)$$

$$\mathrm{FNR} = 1 - \mathrm{Recall} \qquad (20)$$

False Negative Rate (FNR) metrics comparison of the proposed and existing classifiers is illustrated in figure 10 (b). FNR of the suggested AdaBoost-CNN is 12%. But the FNR of the existing CNN, DNN, DBN and MLP classifiers are 16%, 23%, 28% and 30% respectively. It demonstrates that the suggested classifier performs better in terms of FNR than the current methods.

TABLE II. PERFORMANCE COMPARISON OF THE SUGGESTED AND CURRENT APPROACHES

Parameters	Proposed Method	CNN-DNN [15]	Random Forest [18]	DFA [19]
Accuracy	95%	90%	85%	83%
Error	5%	10%	13%	11%
Precision	87%	79%	75%	69%
Recall	85%	73%	69%	64%

Table II illustrates the performance of the suggested and current techniques. The accuracy of the suggested technique is 95%. The suggested approach is compared with the current methods mentioned in the literature reviews like CNN-DNN, Random forest and DFA. The effectiveness of the suggested approach is high when compared to the current techniques. As a result, the proposed Adaptive Boosting CNN was the best choice for forecasting rice crop production based on soil nutrient level.

After observing the results of proposed methodology and the comparative analysis with existing approaches, the accuracy of proposed methodology is increased. The dataset used in this research work is having fixed data values. Whenever an agricultural region experiences a good rainfall, soil moisture, NPK and pH value will change. Due to this, fixed data may show less accuracy. In order to increase the accuracy, smart agricultural system can be designed with the help of sensors that can be used to get the values of NPK and pH value of the soil. This system will give a real time data values which will provide more accuracy in the prediction results.

In order to get the NPK value, soil NPK sensor with power of 9v-24v can be used along with the differential pH sensors. Soil NPK sensor such as tensiometers, volumetric sensors or solid state sensors gives the values of Nitrogen, Phosphorus and Potassium values from the soil and predicts the fertility of the soil with the help of Audrino. Also, pH sensors can be divided into three categories namely, combinational sensors called electro chemical sensors that have two electrodes namely, reference electrode used to keep track of stable level of signals and other one is measuring electrode, used to determine the changes in pH value based on the value of first electrode. It is a basic sensor for differential and laboratory sensors. Differential sensors are used to determine soil pH value but it has three electrodes. Two electrodes are same as the combinational one but third one is for grounding purpose to avoid fouling in reference. Third category called as laboratory pH sensor that uses the technology of combinational sensor with the 12mm glass and has some plastic bodies in it.

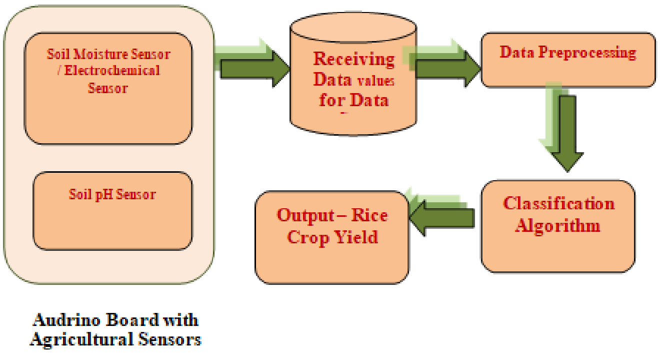

Soil Moisture Sensor / Electrochemical Sensor
Soil pH Sensor
Receiving Data values for Data
Data Preprocessing
Classification Algorithm
Output – Rice Crop Yield
Audrino Board with Agricultural Sensors

Figure 11. Smart Agricultural System for Rice Crop Production

Design of smart agricultural system for rice crop yield has been shown in figure 11. Classification algorithm such as Deep Belief Network, Deep Neural Network, Convolutional Neural Network, Artificial Neural Network etc., along with any optimization algorithm will be used to predict the output yield.

V. CONCLUSION AND FUTURE SCOPE

In this work an adaptive boosting technique was integrated with convolutional neural network called AdaBoost-CNN was proposed for forecasting rice crop yield in the particular soil. Soil nutrients and crop production statistics were taken as the input for the proposed method. This input dataset has some missing or omitted values. In order to fulfill the missing values in the dataset, proposed method utilizes the preprocessing step. Adaptive data cleaning and imputation using most frequent values were used as pre-processing methods. The pre-processed data was then fed to the classifier. The data weights are initialized, and the initial data weight is used to train the first CNN. The weights of the 2nd CNN were then modified utilizing the first CNN. In every iterative process, the classifier affiliated with that repetition was trained utilizing training data also data weights. These action will be perform sequentially for all weak classifiers. Finally the Horse Herd Algorithm was used to find optimal weights of the links in the classifiers in the proposed method. This AdaBoost-CNN classifies the input data and predict the output as six different classes. This proposed classifier compared with existing classifiers such as CNN, DNN, DBN and MLP. The proposed method attains 95% accuracy, 87% precision, 85% recall, 5% error, 96% specificity, 87% F1_Score, 97% NPV and 12% FNR value. Graphical representation prove that the suggested method results are significantly better than those of the current methods. Thus the proposed AdaBoost-CNN classifier was the best choice for forecasting rice crop output based on soil nutrient level. In future, a hybrid approach using a biophysical model and a machine learning approach will be developed to improve the accuracy of crop yield forecasts. Also, based on these results, smart agricultural system for rice crop yield is designed with the help of electro chemical and soil pH sensors. This implementation can be done in future to improve the accuracy of prediction systems with respect to real time values collected from sensors.

REFERENCES

[1] Simkhada, K., & Thapa, R. (2022). Rice Blast, A Major Threat to the Rice Production and its Various Management Techniques. Turkish Journal of Agriculture-Food Science and Technology, 10(2), 147-157.

[2] Abotaleb, M., Ray, S., Mishra, P., Karakaya, K., Shoko, C., Khatib, A. M. G. A., ... & Balloo10, R. Modelling and forecasting of rice production in south Asian countries.

[3] Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. advances in agronomy, 154, 201-255.

[4] Tsujimoto, Y., Rakotoson, T., Tanaka, A., & Saito, K. (2019). Challenges and opportunities for improving N use efficiency for rice production in sub-Saharan Africa. Plant Production Science, 22(4), 413-427.

[5] Shrestha, J., Kandel, M., Subedi, S. and Shah, K.K., 2020. Role of nutrients in rice (Oryza sativa L.): A review. Agrica, 9(1), pp.53-62.

[6] Nishant, P. S., Venkat, P. S., Avinash, B. L., & Jabber, B. (2020, June). Crop yield prediction based on indian agriculture using machine learning. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-4). IEEE.

[7] Li, L., Wang, B., Feng, P., Wang, H., He, Q., Wang, Y., ...& Yu, Q. (2021). Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. Agricultural and Forest Meteorology, 308, 108558.

[8] Hossain, M., Roy, D., Maniruzzaman, M., Biswas, J. C., Naher, U. A., Haque, M., & Kalra, N. (2021). Response of crop water requirement and yield of irrigated rice to elevated temperature in Bangladesh. International Journal of Agronomy, 2021.

[9] Abd-Elmabod, S. K., Muñoz-Rojas, M., Jordán, A., Anaya-Romero, M., Phillips, J. D., Jones, L., ...& de la Rosa, D. (2020). Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. Geoderma, 374, 114453.

[10] Ansari, A., Lin, Y. P., & Lur, H. S. (2021). Evaluating and adapting climate change impacts on rice production in Indonesia: A case study of the Keduang Subwatershed, Central Java. Environments, 8(11), 117.

[11] Grieve, B. D., Duckett, T., Collison, M., Boyd, L., West, J., Yin, H., ...& Pearson, S. (2019). The challenges posed by global broadacre crops in delivering smart agri-robotic solutions: A fundamental rethink is required. Global Food Security, 23, 116-124.

[12] Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation—A review. Information Processing in Agriculture, 7(1), 1-19.

[13] Koklu, M., & Ozkan, I. A. (2020). Multiclass classification of dry beans using computer vision and machine learning

techniques. Computers and Electronics in Agriculture, 174, 105507.

[14] Devi, M., Kumar, J., Malik, D. P., & Mishra, P. (2021). Forecasting of wheat production in Haryana using hybrid time series model. Journal of Agriculture and Food Research, 5, 100175.

[15] Mishra, P., Al Khatib, A. M. G., Sardar, I., Mohammed, J., Karakaya, K., Dash, A., ... & Dubey, A. (2021). Modeling and forecasting of sugarcane production in India. Sugar Tech, 23(6), 1317-1324.

[16] Sharifi, A., 2021. Yield prediction with machine learning algorithms and satellite images. Journal of the Science of Food and Agriculture, 101(3), pp.891-896.

[17] Panigrahi, K.P., Das, H., Sahoo, A.K. and Moharana, S.C., 2020. Maize leaf disease detection and classification using machine learning algorithms. In Progress in Computing, Analytics and Networking (pp. 659-669). Springer, Singapore.

[18] Varma, M., Singh, K. N., & Lama, A. (2022). Exploring the suitability of machine learning algorithms for crop yield forecasting using weather variables.

[19] Ajithkumar, B. (2021). Rice yield forecasting using principal component regression and composite weather variables. Journal of Pharmacognosy and Phytochemistry, 10(2), 595-600.

[20] Jeong, S., Ko, J., & Yeom, J. M. (2022). Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. Science of the Total Environment, 802, 149726.

[21] Chu, L., Jiang, C., Wang, T., Li, Z., &Cai, C. (2021). Mapping and forecasting of rice cropping systems in central China using multiple data sources and phenology-based time-series similarity measurement. Advances in Space Research, 68(9), 3594-3609.

[22] Oikonomidis, A., Catal, C., &Kassahun, A. (2022). Hybrid Deep Learning-based Models for Crop Yield Prediction. Applied artificial intelligence, 1-18.

[23] Mishra, P., Sahu, P. K., Devi, M., Fatih, C., & Williams, A. J. (2021). Forecasting of Rice Production using the Meteorological Factor in Major States in India and its Role in Food Security. International Journal of Agriculture, Environment and Biotechnology, 14(1), 51-62.

[24] Nazir, A., Ullah, S., Saqib, Z. A., Abbas, A., Ali, A., Iqbal, M. S., ...& Butt, M. U. (2021). Estimation and Forecasting of Rice Yield Using Phenology-Based Algorithm and Linear Regression Model on Sentinel-II Satellite Data. Agriculture, 11(10), 1026.

[25] Kandan, M., Niharika, G. S., Lakshmi, M. J., Manikanta, K., &Bhavith, K. (2021, November). Implementation of Crop Yield Forecasting System based on Climatic and Agricultural Parameters. In 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT) (pp. 207-211). IEEE.

[26] Kumar, J., Devi, M., Verma, D., Malik, D. P., & Sharma, A. (2021). Pre-harvest forecast of rice yield based on meteorological parameters using discriminant function analysis. Journal of Agriculture and Food Research, 5, 100194.

[27] Islam, T., Chisty, T. A., & Chakrabarty, A. (2018, December). A deep neural network approach for crop selection and yield prediction in Bangladesh. In 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1-6). IEEE.

[28] Guo, Y., Xiang, H., Li, Z., Ma, F., & Du, C. (2021). Prediction of rice yield in East China based on climate and agronomic traits data using artificial neural networks and partial least squares regression. Agronomy, 11(2), 282.

[29] Bondre, D. A., & Mahagaonkar, S. (2019). Prediction of crop yield and fertilizer recommendation using machine learning algorithms. International Journal of Engineering Applied Sciences and Technology, 4(5), 371-376.

[30] Iizumi, T., Shin, Y., Kim, W., Kim, M., & Choi, J. (2018). Global crop yield forecasting using seasonal climate information from a multi-model ensemble. Climate Services, 11, 13-23.

[31] Pagani, V., Guarneri, T., Busetto, L., Ranghetti, L., Boschetti, M., Movedi, E., ...& Confalonieri, R. (2019). A high-resolution, integrated system for rice yield forecasting at district level. Agricultural Systems, 168, 181-190.

[32] Kour, S., Shitap, M. S., Pradhan, U. K., Paul, R. K., Arya, P., & Kumar, A. (2018). Forecasting of rice yield based on weather parameters in Kheda district of Gujarat, India. International Journal of Agricultural and Statistical Sciences, 14(2), 611-615.

[33] https://www.soilhealth.dac.gov.in/NewHomePage/NutriReport

[34] https://aps.dac.gov.in/APY/Index.htm

[35] Sun, D., Wu, J., Yang, J., & Wu, H. (2021). Intelligent Data Collaboration in Heterogeneous-device IoT Platforms. ACM Transactions on Sensor Networks (TOSN), 17(3), 1-17.

[36] Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., &Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: an experimental review. Journal of Big Data, 7(1), 1-47.

[37] González, S., García, S., Del Ser, J., Rokach, L., & Herrera, F. (2020). A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. Information Fusion, 64, 205-237.

[38] Xu, X., Duan, H., Guo, Y., & Deng, Y. (2020). A cascade adaboost and CNN algorithm for drogue detection in UAV autonomous aerial refueling. Neurocomputing, 408, 121-134.

[39] MiarNaeimi, F., Azizyan, G., & Rashki, M. (2021). Horse herd optimization algorithm: a nature-inspired algorithm for high-dimensional optimization problems. Knowledge-Based Systems, 213, 106711.

[40] Taherkhani, A., Cosma, G., &McGinnity, T. M. (2020). AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning. Neurocomputing, 404, 351-366.