IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 March 2023, accepted 22 April 2023, date of publication 28 April 2023, date of current version 4 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3271410

RESEARCH ARTICLE

A Multimodal Data Fusion and Deep Neural Networks Based Technique for Tea Yield Estimation in Pakistan Using Satellite Imagery

ZEESHAN RAMZAN[1], **H. M. SHAHZAD ASIF**[1], **IRFAN YOUSUF**[1], **AND MUHAMMAD SHAHBAZ**[2]
[1]Department of Computer Science, New Campus, University of Engineering & Technology Lahore, Lahore 54000, Pakistan
[2]Department of Computer Engineering, University of Engineering & Technology Lahore, Lahore 54000, Pakistan

Corresponding author: Zeeshan Ramzan (zramzan@uet.edu.pk)

This work was supported by the Higher Education Commission (HEC) Pakistan as National Research Program for Universities (NRPU) under Grant NRPU 9201.

ABSTRACT Achieving food security has become a major challenge for society. Crop yield estimation is essential for crop monitoring to ensure food security. Manual crop yield estimation is cumbersome and inaccurate and becomes infeasible when scaled up. Machine learning algorithms trained using remotely sensed data have played a vital role in estimating the yield of different crops. Furthermore, to enrich the data provided to a machine learning algorithm, multiple modalities can be combined to improve the predictive performance of these algorithms. In this research, we propose to combine data from multiple modalities, i.e., agrometeorological and remote sensing data, to predict the tea yield at the farm level. The dataset employed in this study is acquired from tea fields of the National Tea and High-Value Crop Research Institute (NTHRI), Mansehra, Pakistan. The remote sensing data of the Landsat-8 satellite is converted to farm-level NDVI statistics through geocoding. Before being used for regression modeling, the final dataset is subjected to some further preprocessing steps, including the selection of features and the optimization of feature sets. This preprocessed data is used to train the three classes of machine learning regression algorithms. Conventional regression algorithms, including Decision Trees, Multilayer Perceptron (MLP), Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Multiple Linear Regression applied with and without interaction terms and stepwise feature inclusion with various kernels. Moreover, the following three variants of the ensemble learning methods have also been applied: random forest, gradient boosting, and XgBoost. Finally, this study proposed a neural architecture for tea yield estimation using Landsat imagery. This deep neural network is built using neural architecture search via Bayesian optimization and have three hidden layers, which can perform complex non-linear modeling. Experimental evaluation is performed through 10-fold cross-validation, and the proposed Deep neural network regression model provided the best predictive performance. The model provided a coefficient of determination (R-squared) of 0.99 with a Mean Square Error (MSE) of 108.17 kg/ha, Root Mean Square Error (RMSE) of 10.87 kg/ha, Mean Absolute Error (MAE) of 2.26 kg/ha and Mean Absolute Percentage Error (MAPE) of 2.92.

INDEX TERMS Multimodal, data fusion, deep neural networks, crop yield forecasting tea yield, Pakistan, satellite imagery.

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

I. INTRODUCTION

Farming has an important role in the advancement of human civilization. The amount of land that is suitable for crop production is shrinking as the world's population continues to increase, yet the demand for food supplies is growing. In the sector of agriculture, many forms of automation are now being developed to meet the increased demand for food. To achieve their goal of increasing crop productivity, agronomists are implementing a variety of innovative practices. Fertilizers, weed management, and insect control methods are being used to combat soil insufficiency and other yield-reducing variables. Furthermore, the examination of soil and weather conditions assists agronomists in better understanding the fields and selecting the correct crop for a certain location based on soil and weather qualities [23].

Considering the scope of agricultural production, since the allocation of land, water, and fertilizers used by farmers cannot be increased beyond a specific limit, improvement in technical efficiency and bringing technological change are the only key to attaining further growth in the agriculture sector. The application of Information Technology (IT) in the field of agriculture has become relatively easier as a result of developments in related technologies [6]. One such advancement is known as Artificial Intelligence (AI), and it is currently being used to make crop yield predictions in advance of the harvest season. The use of AI allows for the identification of crop diseases, the estimation of yield, the making of informed decisions, and so on. Because more sophisticated and high-resolution sensors are becoming available, remote sensing is becoming an increasingly valuable tool. These sensors are used for crop evaluation and yield estimation in the context of environmental, agronomic, or hydroclimatic factors. These AI-based algorithms use the data from remote sensing to model the spread of crop disease based on the agrometeorological parameters for a particular region.

It is essential to perform crop monitoring and estimate the production to control the food supply. Recognizing the harvest conditions and the potential yield in a timely and accurate manner may provide us with the knowledge that is useful in enabling us in taking preventative actions. Several variables are thought to affect yield, thus focusing too much on just one of them could negatively impact yield. Researchers have developed a variety of models that can be used to monitor crops in a variety of topographical areas and for a variety of crops. Variables about the environment and the soil play the most important role in crop monitoring. In addition, the pre-harvest estimate of crop production has a direct influence on the financial aspects of any country and is an extremely important factor in the administration of food supplies [5], [45].

The estimation of tea yield is the main focus of this study as, in comparison to wheat, rice, maze, and other seasonal crops, it has been relatively less explored. In addition, tea is one of the most important imported commodities, and any progress made toward increasing tea production might be to the advantage of both domestic farmers and the nation as a whole. Tea is believed to have originated in the southeast part of China. It is cultivated in areas that experience hot and humid summers and cold and dry winters. The annual amount of precipitation that tea plants are supposed to receive ranges from 1150 to 1400 millimeters, but the precipitation that occurs throughout the year has a greater influence on the yield. Temperature is the most important factor to be considered among all of the environmental parameters since the growth of the tea plant slows or even comes to a complete halt at temperatures that are much higher or lower. In our study, the ecological temperature remains high in Shinkiyari Tea Farms, but its impact on growth is less severe. Additionally, an increase in humidity plays a part in the growth and production of the tea crop. The plants are negatively impacted by the fast wind and hail, which in turn affects the yield. Laycock et al. [39] described the best developing time for tea as having warm and humid days with long daylight periods and average precipitation ideally throughout the evening.

The following are the contributions of our research: In this study, we proposed to combine data from multiple modalities, i.e., agrometeorological and remote sensing data, to predict the tea yield at the farm level. The data employed in this study include yield and agrometeorological parameters that were recorded via various sensors for six years. In addition, data from remote sensing is collected to calculate vegetation indices, which, in conjunction with agrometeorological information, are utilized in the process of model building for yield estimation. The collected dataset is preprocessed as we performed the agrometeorological parameters' correlation with each other and with the yield, which may be used to determine how relevant each of these attributes are in the estimation of yield. We also selected the important features using the ReliefF algorithm to minimize the computational complexity, avoid the curse of dimensionality, and simplify the model to decrease the occurrence of overfitting. This preprocessed data is used to train the three classes of machine learning regression algorithms. Conventional regression algorithms, including Decision Trees, Multilayer Perceptron (MLP), Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Multiple Linear Regression applied with and without interaction terms and stepwise feature inclusion with various kernels. Moreover, the following three variants of the ensemble learning methods have also been explored: random forest, gradient boosting, and XgBoost. Finally, this study proposed a neural architecture for tea yield estimation using data from multiple modalities. This deep neural network is built using neural architecture search using Bayesian Optimization with three hidden layers, which can perform complex non-linear modeling. This proposed deep neural network is an improved technique to estimate tea yield at the farm level. A Neural Architecture Search (NAS) has also been implemented in this study to select the best configuration for a neural network. NAS is performed by exploring the design in three aspects: The search space focuses on limiting the types of neural network architectures that can be developed and optimized. The search strategy specifies how the search space is investigated. The performance estimation method evaluates the capability of a neural network.

Some of the highlights of the study are provided below:

- To predict the tea yield using data from multiple modalities and machine learning algorithms at the farm level in Pakistan.
- To study the agrometeorological parameters' that are affecting the tea yield.
- To study how deep neural networks can be used for tea yield prediction and how they are providing good results with less error as compared to machine learning regression algorithms.
- To implement a Neural Architecture Search (NAS) to select the best configuration for the Neural Networks.

II. RELATED WORK

Crop yield estimation is essential to ensuring food security, therefore various efforts are being made by agriculturists, remote sensing, data mining, and, most notably, information technology researchers. The research objective is broad, so discussing all relevant literature is not useful. However, to highlight the research gap, recent literature on the use of machine learning for crop yield estimation is briefly discussed. Although the context of the proposed approach focuses on tea yield estimation, there is relatively less literature published in this area. Furthermore, because the techniques developed for seasonal crop yield estimation are similar in some ways to tea yield estimation, the related work also provides some notable approaches targeted to crop yield estimation.

A. REMOTE SENSING-BASED APPROACHES

The use of remote sensing allows for an assessment of the temporal and spatial variability of crop yield, in addition to other elements of the agricultural dynamics [47], [63]. It has been shown that the visible band of the electromagnetic spectrum (blue, green, and red) and the near-infrared bands help acquire information on crop type, crop health, soil moisture, nitrogen stress, and crop yield [1], [8], [14], [24], [25], [42], [44], [47], [62]. Improved remote sensing technology has made multispectral images a significant resource for analyzing and monitoring plant health, detecting regions of agricultural stress, and estimation yields. According to Liu et al., [43], data obtained by remote sensing indicated previously unknown geographical and temporal land surface features. Among them were the environmental factors that influenced the growth of crops. It has been discovered in several studies [19], [22], [43], [55] that vegetation indices that were derived via the use of remote sensing techniques have a significant potential for association with crop production and biomass. Agricultural yield studies conducted on a regional scale may provide more extensive insights into crop canopy conditions and yield estimates by making use of satellite images with a coarse or low-resolution resolution.

In the process of crop prediction, many agronomic variables, including vigor, maturity, density, and disease, are often used as yield indicators. However, the phenology, crop health, and stage type affect the spectral reflectance of the crop; to provide an appropriate assessment of plant health using remote sensing. Recent studies [7], [8], [17], [21], [59], [63], [67] have focused on the normalized difference vegetation index, also known as the NDVI, to enhance precision farming. According to research that was done on the monitoring of plant life, there is a correlation between the NDVI and both the leaf area index and the photosynthetic activity of crops. Through the use of the percentage of photosynthetically active light that is absorbed, the NDVI can quantify crop growth in an indirect manner [46], [53].

In a similar study, Baez-Gonzalez et al. [8] used data from Landsat Enhanced Thematic Mapper (ETM) in conjunction with an NDVI model to estimate corn yield. The authors discovered that the estimation had a prediction error of 9.2%. According to Yang et al., [68], the EPIC model that is maintained by the United States Department of Agriculture was used to forecast yield, and there was an error of less than 10% between the predicted yield and the actual yield. Images captured by NOAA's advanced high-resolution radiometer were utilized to generate an NDVI, which Baez-Gonzalez et al. [7] then used to model a corn yield.

Gopalapillai and Tian [21] predicted corn yield from nine fields across two years of data and found that the correlation values varied from 0.13 to 0.98. The authors obtained Landsat images of the corn fields and used the NDVI values to construct estimations to forecast yield. After looking at all nine fields, the authors discovered that NDVI and yield had a correlation coefficient of 0.54 with one another. It has been demonstrated that the soil-adjusted vegetation index has a bias toward decreasing soil brightness. This is a problem that has been explored by researchers such as Miura et al. [49] and Lamb et al. [38]. After researching potato yield estimates with soil-adjusted vegetation index derived from high-resolution aerial multispectral data, Jayanthi [31] developed a variety of yield estimate models. To take into consideration the first-order optical interactions that occur between the ground and plants, Huete et al. [28] included a calibration term for the soil into the NDVI equation. Using ground truth data from 50 farms and TERRA MODIS images, Bala et al. [9] estimated a potato yield with an error of 15.

B. NEURAL NETWORKS-BASED APPROACHES

To estimate yields using a variety of modalities including remote sensing, weather, soil properties, agro-management techniques, and others, machine learning, and statistical learning approaches are being used. However, it is essential to keep in mind that sophisticated techniques for predictive modeling, such as artificial neural networks, may be used to deal with such data to tackle difficult problems. Several groups of researchers have taken on the challenge of estimating agricultural yields by employing multilayer perceptron [18], deep neural network [37], [40], convolutional neural networks [50], recurrent neural networks or Long

Short-term Memory [20] and a combination of more than one approach [61].

Nevavuori et al. [50] used convolutional neural networks to forecast crop yields, as well as identify crop types and weeds and estimate biomass. The authors claim that the failure of farm owners to acquire yield mapping devices has forced them to rely on remote sensing and images collected by UAVs to achieve the stated goals. Convolutional neural networks are used because of their incredible performance on visual classification problems. The authors have performed learning using multispectral data such as RGB and near-infrared imagery, either directly or by calculating vegetation indices such as NDVI from this data. To perform learning, the authors used a six-layer deep convolutional neural network with hyperparameter optimization. The Adadelta optimizer is used to perform the learning, and the mean absolute error is used as the loss function. The research reveals that RGB data is superior to vegetation indices for yield estimation modeling. Furthermore, using late-season imagery, the authors reported a mean absolute percentage error of 12.6%.

In a similar attempt, Khan et al. [35] also used neural networks for yield prediction, evapotranspiration evaluation, and biomass estimate utilizing images received from the OLI sensor aboard the Landsat 8 satellite. Kumar et al. [36] trained neural networks, random forests, support vector regression, and linear regression using multi-spectral data with a resolution of 1 meter. These indices were utilized to measure different types of vegetation indices. The model carried out an analysis using winter wheat, made a prediction regarding crop yield based on different growing seasons, and reached an r-squared value of 0.95. To accomplish leaf area index estimate from the simulation data of multi-spectral imagery, Liang et al. [41] have employed curve fitting, random forest, and artificial neural network. Based on the results of the predictive study, it was determined that the proposed approach had an r-squared value of 0.98. The three crops for which crop yield forecasts were made by Johnson et al. [32] in the Canadian Prairies were spring wheat, canola, and barley. MODIS and NOAA NDVI and EVI data were used in the predictive modeling process. The performance of three potential models for yield estimates using Bayesian neural networks, model-based recursive partitioning, and multiple linear regression was compared using Bayesian neural networks, model-based recursive partitioning, and multiple linear regression. The results indicated that, except for barley, there was no significant performance disparity between the three models.

Deep neural networks were used by Khaki et al. [34] to make predictions about the yields of maize and soya beans. To create accurate forecasts of agricultural production, the authors recommend using a CNN-RNN architecture as well as a deep fully-connected neural network, Lasso, and random forest. The authors have stated that they were able to get an r-squared value of 0.95 by employing the CNN-RNN model which is higher as compared to other models.

Similarly, Schwalbert et al. [58] published a crop yield prediction model for the soya bean crop using a deep neural network in addition to two other methods. The authors have employed techniques such as random forest, linear regression, and long short-term memory-based networks to create accurate forecasts utilizing data from weather and remote sensing. For crop yield prediction, the study made use of EVI and NDVI indices, as well as precipitation and surface temperature as independent variables, and also experimented with data gathered during different growing seasons.

C. TEA YIELD ESTIMATION

The estimation of tea yield is a particular case of agricultural yield estimation, with one significant distinction being that it is a perennial crop rather than a seasonal one. To provide an overview of the literature that has contributed to the yield estimation of tea crops, some related studies are discussed. Jui et al. [33] presented a unique strategy for estimating the productivity of tea crops in their study. The research made use of hydro-meteorological data acquired from 20 locations in Bangladesh from 1981 to 2020 using satellite technology. Support vector machines and a dragonfly optimization were used to select the optimal feature set for regression modeling. A hybrid spatiotemporal random forest model is used to forecast the yield of the tea crop. The authors obtained an r-squared value of 0.97 when using their proposed method for the dataset containing tea crop data. Similarly, Dhekale et al. [13] have focused their efforts on predicting the tea yield in Bangladesh by utilizing time series data. The authors carried out modeling utilizing time series estimation techniques, in addition to doing research into the growth and trend behavior of tea production. Mila et al. [48] also targeted the case of tea yield estimation in Bangladesh and incorporated the use of time series data to make estimations using various settings of ARIMA.

Phan et al. [52] have proposed a technique that is based on spatiotemporal remote sensing data to carry out tea yield estimation. The authors have used the data to compute NDVI from the multispectral data, and the authors have shown that NDVI has a high r-squared value of 0.79 with mean temperature. Utilizing this information to train three different types of regression algorithms allows for the accurate prediction of yields. In contrast to multiple linear regression and random forest, the model that was based on support vector regression offered a greater predictive performance of 0.87 for r-squared.

Tea yield estimation on a regional scale has been carried out by Raj et al. [54] using agrometeorological data. In the course of the research, four different regression algorithms were tested through their experiments. The authors reported that simple regression models, such as stepwise linear regression and autoregressive integrated moving average regression, were not successful in performing accurate predictions of tea yield. Specifically, the authors focused on how these models failed to accurately predict tea yield. On the other hand, being non-linear models, artificial neural networks and vector autoregressive models have shown greater predictive performance. Additionally, temperature is the most important factor in determining the level of variability in tea yield, while rainfall is the second most important factor.

Tea yield forecasts were made by Ahmed et al. [5] using agrometeorological and soil parameters. The tea research institute provided the data that was utilized for the study, which included farm-level yields to train the regression algorithms. The authors employed feature pre-processing techniques such as outlier removal, feature transformation, and feature selection to generate the final feature set to perform regression modeling. The authors employed seven regression techniques to generate predictions before proposing an ensemble built from multilayer perceptrons as weak learners that were optimally combined using a novel strategy. The proposed ensemble of neural networks produced the highest r-squared value, followed by Gaussian process regression and random forest.

Batool et al. [10] conducted yield forecasts based on their research using data obtained from soil sources, agrometeorological sources, and remote sensing. The authors were able to accomplish yield prediction with the help of the crop simulation model AquaCrop by conducting tuning of the model parameters. In addition, the authors have performed regression learning using a variety of statistical and machine learning approaches and have experimented with various approaches. When compared to other regression models and the crop simulation model, it has been observed that XGBoost provides the highest performance when it comes to yield prediction. Furthermore, the regression models outperformed the crop simulation model since data for tuning all AquaCrop parameters for tea crop yield prediction was not available. Furthermore, the authors argued that providing the necessary data and further optimizing model parameters might result in increased performance by AquaCrop.

III. MATERIALS AND METHODS

This section describes the study area, followed by a description of data sources and data preparation steps. Cross-validation was chosen over holdout validation because it is a more rigorous method of evaluation and helps to eliminate the possibility of train-test bias when performing model building and evaluation. The modeling process begins by collecting data from three different sources and processing it appropriately. Conventional regression modeling is used in the process of building the candidate models. By using the proposed evaluation method, seven different regression algorithms from the conventional regression category are trained and evaluated. The deep neural networks-based regression model that was proposed (through neural architecture search) is trained on a dataset that was processed differently, and its performance is evaluated using the same model evaluation strategy that was used to evaluate other candidate models. A depiction of the data preparation, model building, and evaluation processes can be found in Figure 1.

A. STUDY AREA

The study is carried out at tea farms in Shinkiyari, Mansehra, Pakistan (latitude, 34°27’ E; longitude, 73°16’ N; elevation, 542m) as shown in Figure 2. These farms span an area of 50 acres of land with 24 acres of tea farms and are operated and managed by the National Tea and High-Value Crop Research Institute (NTHRI) of Pakistan. These tea farms are part of a research facility that specializes in tea farming, processing, and promotion at a national level. The facility is situated in a climate range that is suitable for the cultivation of tea plants and is located in a region that is declared suitable for the cultivation of tea plants. Because of these features and the availability of research staff, it is feasible to collect precise data about the parameters of the weather, the soil, and the plant’s health, as well as the yield obtained at each round of plucking. The data that is analyzed and utilized in the development of the prediction model spans the period from 2015 to 2021, and the final dataset includes a total of 1080 instances of yield along with other attributes.

B. DATA SOURCES

The data includes yield and agrometeorological parameters that were recorded every month over six years. In addition, data from remote sensing is collected to calculate vegetation indices, which, in conjunction with agrometeorological information, are utilized in the process of model building for yield estimation.

1) LANDSAT-8 DATA

In February 2013, the United States deployed a remote-sensing satellite named Landsat 8. This satellite has two sensors: Thermal Infrared Sensor (TIRS), and Operational Land Imager (OLI), which capture data in several wavelength ranges. These sensors have a spatial resolution of 100 meters for thermal, 30 meters for visible, near-infrared, and short-wave infrared, and 15 meters for panchromatic bands. NASA and the U.S. Geological Survey (USGS) jointly developed the Earth Resources Observation and Science (EROS) center, which stores the data from the Landsat 8 satellite. Before the launch of Landsat-8, Landsat 7’s Enhanced Thematic Mapper Plus (ETM+) band, which EROS also supplies, was used to collect multispectral data. ArcGIS Pro 3.0.1 analyzes multispectral images acquired using the OLI of Landsat-8 or the ETM+ of Landsat 7. Normalized Difference Vegetation Index is computed using images captured throughout the growing cycle at precise coordinate points determined using professional GPS equipment. Compared to other vegetation indices, the Normalized Difference Vegetation Index (NDVI) is the most suitable vegetation index for assessing plant health and, ultimately, for estimating yield.

2) IN SITU SOIL DATA

Warm, humid weather and deep, fertile, well-drained soil made of laterite that is rich in organic matter are perfect for cultivating tea. Laterite soils are very smooth while they are wet, but when they are dry, they become lumpy and difficult to grow. Laterite soils may be found on high, flat erosion surfaces in locations that get moderate and seasonal rainfall. Tea plants need constant warmth and humidity without the possibility of frost to be healthy and productive. The growth of uniform leaves throughout the year is encouraged by precipitation that is both steady and light. In contrast to the majority of crops, tea plants need acidic soil with a PH level ranging from 4.5 to 5.5 to flourish. There is a chance that tea plants growing in soil with a PH level of 7.0 may not develop properly or will experience acute malnutrition. Therefore, the PH level is measured at a depth of 0.3 m to be used as an attribute. In comparison to other attributes, the soil PH level measurement was a manual process, and some inconsistencies in the form of missing and noisy readings were observed and the attribute was discarded before model building. However, it is advised that we incorporate PH level as an attribute, which we will do in our future work.

3) WEATHER DATA

The weather data is required as part of agrometeorological data and is acquired using an onsite weather station HOBO RX3000 which provides site-specific environmental data through a web-based interface. The model HOBO RX3000 is specifically tailored for environmental research, crop management, and greenhouse operations and performs data logging at a rate of once per second to once every 18 hours. It is a very versatile and configurable weather station that provides environmental data for barometric pressure, air velocity, carbon dioxide, wind, differential pressure, dew point, leaf wetness, evapotranspiration, light intensity, relative humidity, rainfall, soil temperature, soil moisture, solar radiation and atmospheric temperature among others. Before the installation of HOBO RX3000 in 2017, the data was recorded by onsite staff or using WH2950 solar weather station wirelessly. This weather station provides data logging for temperature, humidity, wind speed, wind direction, rainfall, dew point, heat index, solar radiation, and barometric pressure.

C. DATA ANALYSIS AND PREPARATION

The results of a Pearson correlation analysis between each pair of variables are presented in Figure 3. This kind of analysis is used to determine whether or not the relationship between a variable and other variables is linear. The average temperature, the highest temperature, and the lowest temperature all have a very close relationship with one another. Because of the strong dependence that these factors have on one another, there is a positive linear relationship that can be drawn between the three temperature values. There is a negative linear relationship between the amount of rainfall and the PH level, which indicates that an increase in one variable will result in a decrease in the other variable. A negative link is shown by the fact that more rainfall causes a decrease in the PH level. There is a slightly positive relationship between rainfall and humidity because rainfall causes a rise in the level of humidity, which is often lower in the case that there is wind. In addition, the fact that there is a positive relationship between the two variables (rainfall and humidity) explains why there is a negative relationship between the PH and the humidity when there is rainfall.

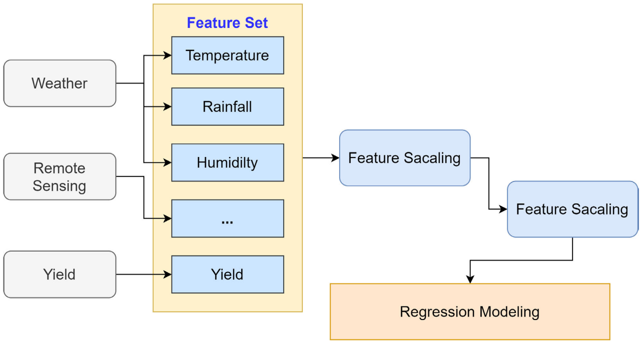


FIGURE 1. Proposed regression modeling and evaluation framework.

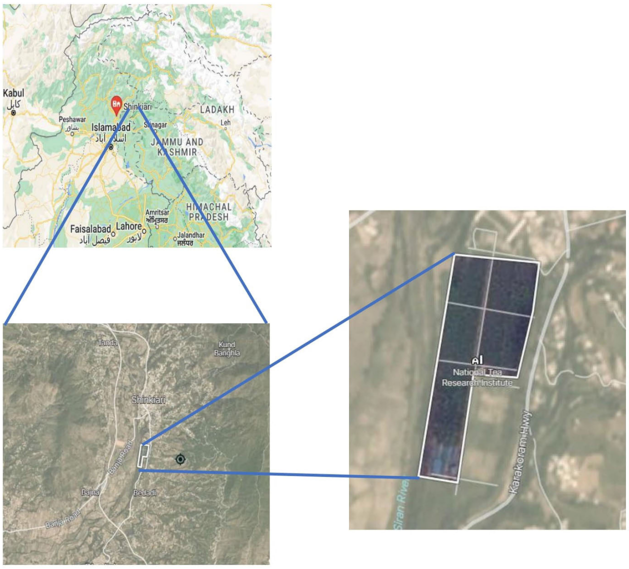
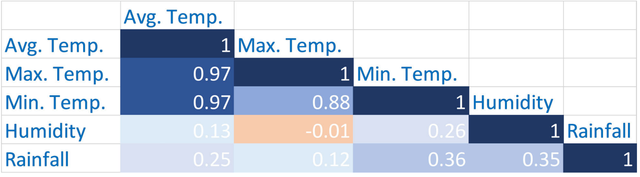


FIGURE 2. Pin of coordinate points (34.45°E, 73.27°N) at the map.



	Avg. Temp.				
Avg. Temp.	1	Max. Temp.			
Max. Temp.	0.97	1	Min. Temp.		
Min. Temp.	0.97	0.88	1	Humidity	
Humidity	0.13	-0.01	0.26	1	Rainfall
Rainfall	0.25	0.12	0.36	0.35	1

FIGURE 3. Correlation between various agrometeorological parameters.

The agrometeorological parameters' correlation with the yield is shown in Figure 4, which may be used to determine how relevant each of these attributes is in the estimation of yield. Rainfall has the highest correlation with the yield indicating a higher rain in the region result in increased tea growth and yield. Three temperature attributes minimum, maximum, and average temperatures are positively correlated with the yield with minimum temperature having the highest relation.

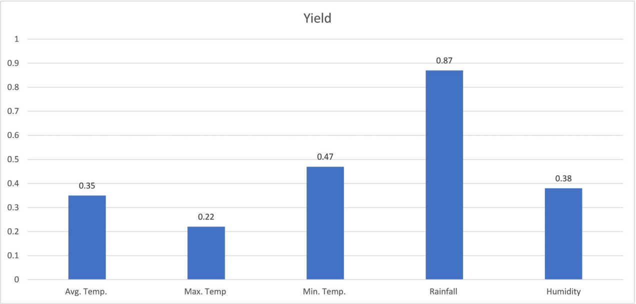


FIGURE 4. Correlation of various agrometeorological parameters with yield.

The fact that a severe drop in temperature slows the growth and yield of tea and therefore a positive change in minimum temperature results in improved yield. The humidity also represents a positive relationship with yield with a relatively lower value of correlation which indicates that it affects the amount of tea growth but the variation in humidity in the region doesn't significantly affect the growth and yield of tea. Furthermore, all of the factors show a positive relationship with the yield that is realistic and suggests that these parameters might be helpful when intending to estimate tea yield.

1) FEATURE SCALING

Numerous machine learning problems employ scaling techniques to improve the feature set for a variety of reasons. It is essential to scale the range of characteristics to a preset range because various factors are measured in different units and may have a large range. It is required to scale the features due to the possibility that the objective function may not execute as per requirement. Regardless of the ranges of the other feature characteristics, the distance measurement is dependent on the feature with the longest range, making this a critical factor for distance-based algorithms. Since each feature contributes about the same amount to the overall distance measurement, it is ideal to bring their respective ranges into parity. Feature normalization techniques have the potential to enhance the convergence characteristic of gradient descent and stochastic gradient descent in some circumstances. The time required for selecting support vectors in an SVM model may be sped up by scaling the features, but the model's accuracy would suffer as a result. Min-max scaling, mean normalization, scaling to unit length, and standardization are well-liked techniques for altering the size of features. We have used neural networks, support vector machines, perceptron, and radial basis function networks due to the advantages of standardization in these approaches. The equation given below is used to calculate the data with a zero-mean and unit standard deviation distribution as a consequence of normalization.

$$x' = \frac{x - \overline{x}}{\sigma} \tag{1}$$

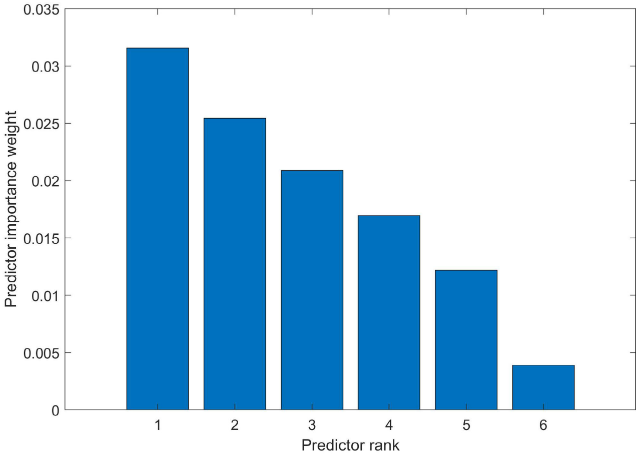


FIGURE 5. Feature ranking performed using ReliefF algorithm.

2) FEATURE SELECTION

Adding supplementary characteristics to the model may enhance predictions, but providing redundant or unnecessary information may lead the prediction process to become confused or add complexity to the computation. Features with substantial correlations to one another or the capacity of one parameter to fulfill the task of another are deemed unnecessary, but those with no relationship to yield are dismissed. Consequently, the selection of features may minimize computational complexity, help avoid the curse of dimensionality, and simplify the model. As an advantage, it may increase the accuracy of forecasts by decreasing the occurrence of overfitting.

Feature selection is a commonly used process during machine learning-based tasks and hence plenty of approaches are available. All these approaches are equally useful and improve the performance of the model but their applicability depends on the nature of the task. However, following the available research guidelines, it is easy to find a suitable feature selection approach for any machine learning task. In this study, a combination of two techniques [3], [4] is used which includes ReliefF [57] and sequential feature selection approaches [2]. The ReliefF algorithm encourages features that assign separate values to neighbors in different classes, while it discourages features that assign distinct values to neighbors in the same class. To repeat, the algorithm favors features that assign different values to neighboring classes. The algorithm also shows the method's weighting structure, in addition to stating the attributes and their respective relevance. Positively weighted features are more likely to lead to accurate predictions, but negatively weighted features have the reverse impact. The y-axes in Figure 5 indicate the relative relevance of the individual traits, whilst the x-axes indicate their scores. As seen in the Figure, all the attributes have positive values. Forward inclusion of features is performed to add one feature at a time to see a reduction in RMSE until convergence.

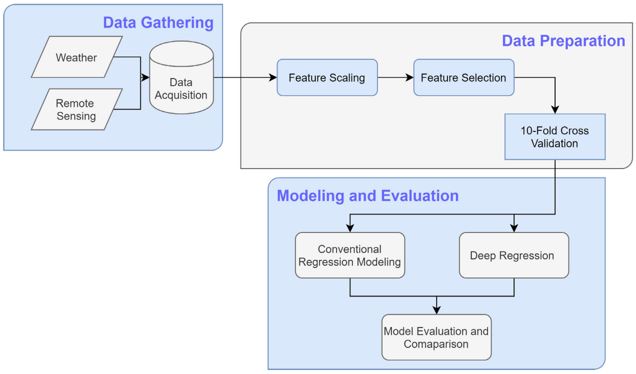


FIGURE 6. Framework for feature set preparation for modeling of conventional regression algorithms.

D. REGRESSION MODELING

Regression is one of the most widely used methods for estimation problems and therefore performing regression modeling for yield estimation makes sense. To perform the prediction of crop (tea) yield using remote sensing and agrometeorological data several regression algorithms are considered. After careful assessment and relevance, seven of these regression algorithms are chosen to perform modeling of tea yield estimation using the processed dataset. The data gathered from multiple modalities is combined by linearizing and concatenating and then processed for regression modeling. Three data pre-processing stages are carried out which prepare the data into a suitable form to perform regression modeling. The general framework of regression modeling is provided in Figure 6 and the regression algorithms used for model building are described briefly in the following sections.

E. DEEP NEURAL NETWORKS

The deep neural network architecture design choices are to be made during the process of designing its architecture, and these choices can have significant impacts on the performance of the produced model. Selecting the best configuration for a neural network's number of layers and the number of neurons in each layer is a difficult task. Not only are the hyperparameters (such as activation function, learning rate, regularization, etc.) crucial, but getting them right can be tricky. There isn't a set of rules that control the process of creating a neural network architecture, unlike the case with many conventional regression techniques. The problem becomes excessively complex as soon as we try to create a deep architecture with more than one hidden layer. This issue can be addressed with the use of a process known as Neural Architecture Search [15], [66] (NAS), which automates the design of the architecture. It's the natural progression of machine learning automation (AutoML [29]) and it shares considerable space with hyperparameter optimization [16] and meta-learning [65]. Image classification [56], [70], object detection [70], and semantic segmentation [12] are only a few examples of the tasks in which NAS approaches have shown superior to manually developed architectures [69], [70] Design-wise, the techniques for NAS may be broken down into categories which are based on three design aspects: (i) search space, (ii) search strategy, and (iii) performance estimation method [15].

TABLE 1. Search space used for neural architecture search.

#	Search Parameter	Range
1	Number of Fully Connected Layers	[1-3]
2	First Layer Size	[1-300]
3	Second Layer Size	[1-300]
4	Third Layer Size	[1-300]
5	Standardize Data	[TRUE, FALSE]
6	Regularization Strength	[1e-5,1e5]
7	Activation	[Sigmoid, Tanh, ReLU, None]
8	Layer Biases Initializer	[Zeros, Ones]
9	Layer Weights Initializer	[Gloroot, He]
#	Output Layer Activation	[Sigmoid, Tanh, ReLU, Linear]

1) What kinds of neural network architectures can be developed and optimized are limited by the search space.
2) The search strategy specifies how the search space is actually investigated.
3) The capability of a neural network is evaluated through the performance estimation method.

1) SEARCH SPACE

The limitations imposed by the search space impose restrictions on the types of theoretical structures that may be modeled. Utilizing prior knowledge of the general characteristics of designs that are appropriate for a certain task enables one to reduce and limit the search space to a more manageable level. Nevertheless, this also adds human bias, which may make it more difficult to uncover novel architectural building blocks that go beyond the confines of current human understanding. To limit the search space to minimize the search time and discovery of excessively complex architecture, we have used a modest search space. Table 1 provides the search parameters and the range for which they are searched.

2) SEARCH STRATEGY

The search space, which might be exponentially large or even unbounded, is what is meant to be explored, and a search strategy details how to do so. This involves the usual exploratory trade-off since it is desired to rapidly uncover high-performing designs. Moreover, it is also undesirable to converge too fast to a region of suboptimal architectures and therefore the search strategy incorporates both of these aspects.

To achieve the objective of the search strategy, we have used 10-fold cross-validation to estimate generalization performance. Moreover, there are various approaches to performing the parameter search with no clear winner. The parameter search can be performed using grid search or random search or using optimization methods. Grid search is an exhaustive method and may require a large amount of time if

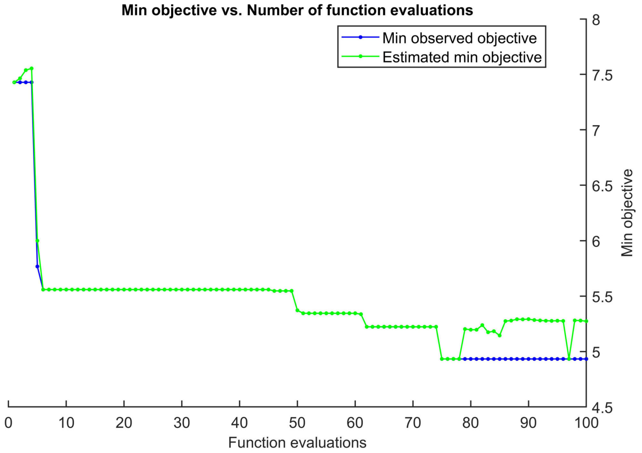


FIGURE 7. Optimization.

TABLE 2. Optimization objectives achieved by Bayesian optimization during parameter search.

#	Search Parameter	Value
1	Number of Objective Evaluations	100
2	Minimum Estimated Objective	5.2739
3	Minimum Objective	4.9328
4	Iterations	1000
5	Training Loss	9.5442
6	Gradient	12.241

the search space is large whereas the random search can be suboptimal. Optimization-based methods include gradient-based optimization, Bayesian optimization, and evolutionary optimization.

To balance the trade-off between search time and optimal parameters, we have opted to use Bayesian optimization as a search strategy. By continually testing different values for the model's hyperparameters, Bayesian optimization seeks to learn as much as possible about the objective function, particularly the optimal configuration. The exploration of the hyperparameters is done with the greatest degree of uncertainty, and the exploitation is done to guarantee the hyperparameters are close to optimal. Bayesian optimization is a global optimization method for black-box functions with noise. By constructing a probabilistic model of the function mapping hyperparameter values to the goal, Bayesian optimization may be used to optimize hyperparameters in the context of a validation set. By reasoning about the quality of experiments beforehand, Bayesian optimization has been demonstrated to outperform grid search and random search as demonstrated by various studies [11], [30], [60], [64]. This is because it requires fewer evaluations to get a satisfactory solution. Figure 7 provides the graph of the minimum estimated objective function value and minimum observed objective value for 100 function evaluations performed using Bayesian optimization. The optimization objectives which were achieved by Bayesian Optimization during parameter search are provided in Table 2. It is to be noted that the 100 optimization evaluation was performed and a minimum observed objective of 4.9328 was achieved.

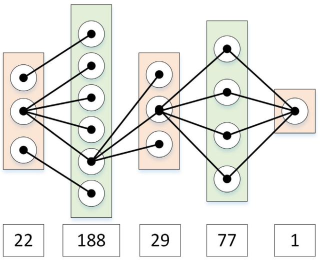


FIGURE 8. Searched deep regression architecture.

TABLE 3. Architecture searched using Bayesian optimization.

#	Search Parameter	Optimized Value
1	Number of Fully Connected Layers	3
2	First Layer Size	188
3	Second Layer Size	29
4	Third Layer Size	77
5	Standardize Data	TRUE
6	Regularization Strength	2.24E-06
7	Activation	ReLU
8	Layer Biases Initializer	Zeros
9	Layer Weights Initializer	He
10	Output Layer Activation	Linear

3) PERFORMANCE ESTIMATION APPROACH

Discovering architectures with superior prediction performance on unknown data is a typical objective in NAS. The quickest and easiest approach to gain a better understanding of the design's performance is to train and validate it on data, however, this technique is computationally expensive and limits the variety of viable topologies. As a result, there has been a lot of recent work devoted to discovering methods to make these performance predictions more reasonable Mean squared error is used as an evaluation measure to evaluate the performance of a specified architecture. This evaluation metric performs computation using 10-fold cross-validation to ensure minimum prediction error as well as generalization.

4) SEARCHED ARCHITECTURE AND TRAINING PARAMETERS

The NAS is performed using the defined search space and search strategy. The evaluation of the searched architectures and hyperparameters is performed using the performance estimation method and the finalized architecture is illustrated in Figure 8 and detailed in Table 3.

Table 4 lists the training parameters utilized after an optimal search during training of the architecture represented in Figure 8 and detailed in Table 3. The LBFGS [51] is a limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm which is a loss function minimization approach in which the algorithm seeks to minimize the MSE.

TABLE 4. Architecture searched using Bayesian optimization.

#	Training Parameter	Value
1	Training Solver	LBFGS
2	Max Iterations	1000
3	Loss Tolerance	1e-6
4	Step Tolerance	1e-6
5	Validation Data	10-fold CV
6	Validation Patience	6

F. IMPLEMENTATION

Keras installed on Python (version 3.7; Google Inc., Mountain View, California, USA) and the Scikit-learn Library (version 0.23.0; Google Inc.) were used for regression modeling. ArcGIS software (version 10.4; was used to process raster data. Python and MATLAB 2021 were used for numerical calculation, modeling, and analysis.

IV. RESULTS AND DISCUSSION

In this part, the defined model evaluation technique and evaluation metrics are used to report experimental outcomes. The evaluation metrics are used to assess the results of several predictive modeling techniques, and the best one is then identified. It is worth noting that the proposed method is quite helpful in modeling the problem, despite having somewhat lower performance compared to the top-performing model because of the limited dataset size. The following part describes the evaluation metrics and approach used to evaluate the models, and then the experimental outcomes are presented for each model.

A. MODEL EVALUATION

Model evaluation is the process of quantification of the performance of a prediction model. This study performs regression modeling and various model evaluation metrics are used by the researchers to evaluate the performance of these models. Before the application of any evaluation metric, it is important to define a model evaluation strategy. Cross-validation and holdout validation are the two most widely used model evaluation methods. As the problem of yield estimation has a small number of instances and fewer features therefore holdout validation is not an appropriate method as it divides the data into a random train-test split. Cross-validation is a statistical procedure in which the data is split into k equal parts before being analyzed. In this study, we have used a 10-fold cross-validation, in which the dataset was split into 10 equal parts, and evaluated the model 10 times. During each cycle, nine subsets of the data are utilized for training, and one subset is used for model validation. A total of 10 iterations are performed, with one set of data serving as an evaluation and the other nine serving as training. To calculate the validity of the cross-validation, we average the results from 10 separate runs of the regression. The approach is a rigorous alternative for holdout validation, and it reduces the potential for bias in data partitioning by repeating the process many times on different subsets of training and validation data.

TABLE 5. Predictive performance of the regression algorithms.

Regression Algorithm	RMSE	R^2	MSE	MAE	MAPE	Inference
Deep Regression (Proposed)	10.87	0.99	108.17	2.26	2.92	29000 *OPS*
Linear Regression	63.96	0.81	4090.60	48.06	12.02	7600 *OPS*
Interactions Linear	45.75	0.90	2093.40	31.42	8.41	7000 *OPS*
Stepwise Linear	45.97	0.90	2113.40	31.64	8.46	9200 *OPS*
Decision Trees	29.06	0.96	844.46	7.03	3.82	32000 *OPS*
SVR (Linear)	97.44	0.57	9493.80	40.19	10.07	29000 *OPS*
SVR (Quadratic)	36.89	0.94	1360.70	19.83	6.23	34000 *OPS*
SVR (Cubic)	30.31	0.96	918.78	15.01	5.32	34000 *OPS*
SVR (Gaussian)	61.61	0.83	3795.40	21.13	6.48	32000 *OPS*
GPR (Rational Quadratic)	12.72	0.99	161.79	2.85	3.03	7100 *OPS*
GPR (Squared Exponential)	13.00	0.99	169.09	2.94	3.05	5800 *OPS*
GPR (Matern 5/2)	11.92	0.99	142.08	2.47	2.96	5300 *OPS*
GPR (Exponential)	13.46	0.99	181.14	3.00	3.06	10000 *OPS*
Gradient Boosting	29.50	0.96	870.25	14.57	5.24	8300 *OPS*
Random Forest	31.03	0.96	962.87	12.20	4.79	6200 *OPS*
XGBoost	13.84	0.99	191.63	4.75	3.39	540 *OPS*

B. EVALUATION METRICS

There are a variety of evaluation metrics that are often used in the process of evaluating the performance of a regression algorithm; however, the mean squared error and the mean absolute error are the evaluation techniques that are the most generally used. The squaring of the error that is calculated between the predicted value and the ground-truth value is the drawback of using mean squared error. This can be avoided by using mean absolute error, which takes the modulus rather than the square, or by using root mean squared error, which takes the square root of the calculated value to avoid the effect of squaring. The appendix provides the description and formulas for the calculation of these three metrics.

C. EXPERIMENTAL RESULTS

The findings of the experiment are provided in terms of RMSE, R^2, MSE, MAE, MAPE, and inference speed. The reported values are the mean of 10-fold cross-validation. Even though the speed at which inferences are made is not a factor in how well a model predicts, they are included to bring attention to the computational complexity of the models. The speed of inference, measured in terms of Observations Per Second (OPS), is monitored on the workstation that is utilized for the experimental assessment. The parameter is calculated by taking the inference time on several examples and averaging it to calculate the number of observations that can be processed per second. This inference speed is only relevant when comparing models that are tested on this specific computer. It is not possible to utilize it to make comparisons between models that were evaluated using different computers.

The predictive performance of the candidate regression algorithms is provided in Table 5. This table demonstrates that the proposed deep neural network regression algorithm exhibits superior predictive performance in comparison to its competitors. The coefficient of determination, also known as R^2, is the most important factor that determines how well a model can perform regression analysis. On the other hand, error measurements like RMSE, MSE, MAE, and MAPE are more open to interpretation. Even though the quantitative scores of each of these measures are distinct from one another, there is a correlation between them. When it comes to the yield estimating challenge, MSE is the most popular and frequently reported measure. On the other hand, the MAPE offers a simple method for understanding the performance of a model by supplying the percentage of absolute error and minimizing the impact of the range of values that are being predicted. It is common practice to consider MAPE in regression to be equivalent to accuracy in classification models.

Multiple modalities, including agrometeorological data and remote sensing images from the Landsat-8 satellite, have been combined to predict tea crop yield. We proposed to derive NDVI from the multispectral data collected by Landsat-8 and to include it with other agrometeorological parameters. Training and evaluation of several different basic and ensemble regression algorithms are undertaken as part of the examination of regression. To execute tea yield prediction, the final model is based on neural architecture search, and it forms a deep neural network with three hidden layers. According to the results of the experimental evaluation, the proposed model is better than alternative methods in terms of r-squared and other evaluation metrics.

The comparison of various candidate models, reported in Table 5 indicate that Gaussian process regression, XGBoost and proposed deep neural networks-based regression model have the highest values of 0.99 for R^2. It is a measure of how much of the variation in yield can be explained by the predictor variables used in these regression models. whereas the strength of the association between an independent variable and a dependent variable may be understood by the correlation, which is presented in Figure 4. When it comes to these candidate models, achieving an R^2 value of 0.99 indicates that nearly 99% of the observed variance can be explained by the model's inputs. It is a quite reasonable value for the coefficient of determination and indicates that any of the three regression algorithms can be used for yield prediction in this scenario. It is however important to emphasize that a high R^2 doesn't always mean a good model and may be the result of a bias in the model. Therefore the further analysis is also carried out to perform residual analysis and response plot.

Calculating the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) allows one to measure the amount of error that exists between the actual values of yield and the expected values. These measurements provide an approximation of how accurately the predicted values of yield are predicting the actual values. The regression model that is based on deep neural networks has provided the lowest values for these error measures. These error measurements can be used to undertake comparisons of the predicted performance produced by various methods. GPR is the second-best-performing algorithm in terms of predictive performance. XGBoost is a close alternative to GPR and comes in third place in terms of MSE, RMSE, MAE, and MAPE. However, the GPR model with Matern 5/2 kernel is the best-performing model after the deep regression model. The measure of error for GPR and XGBoost is slightly higher than the proposed model.

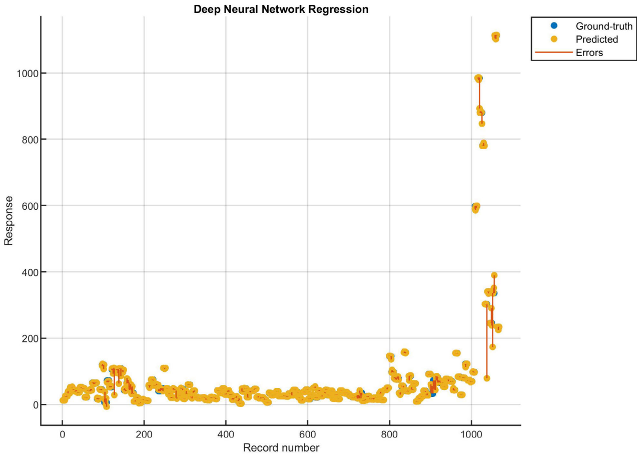


FIGURE 9. Response plot providing ground-truth and predicted values plot.

While inference speed is a major factor in many applications but in our case, it is usually more than adequate for prediction algorithms. The decision tree regression has an inference speed of 32000, whereas the support vector regression with quadratic and cubic kernels has produced the highest inference speed of 34000 OPS. The inference speed of the proposed deep regression model is 29000 OPS, which is fairly reasonable and comparable to the two fastest algorithms. The GPR has an intermediate inference speed that can range anywhere from 5,300 to 10,000 OPS for different kernels, whilst the XGBoost has the slowest inference speed of 540 OPS. Therefore, the proposed deep regression model's predictive performance benefits from the added advantage of inference speed.

The response plot that was generated by the regression model, which in our research was referred to as Deep Regression, is depicted in figure 9. The record number is shown on the x-axis, while the yield is shown on the y-axis of the chart. The yield's ground-truth value, shown by the green circle, is contrasted with the predicted value and is shown by the yellow circle. The errors that were made in the predicted values as compared to the ground-truth values of yield are represented by the vertical red lines in the graph. The plot indicates that the prediction error is reasonably low for the vast majority of the records, even though it is relatively high for a few records. The similarity may be observed in the regression plot of Figure 10, which displays both the predicted and the ground-truth values shown with the regression line with the best fitting. If the vertical gap between the regression line and the observations is modest, the regression model is robust.

The plot showing disparities between the actual values and the values that were predicted plays an important role in regression analysis. This plot is known as the residual plot.

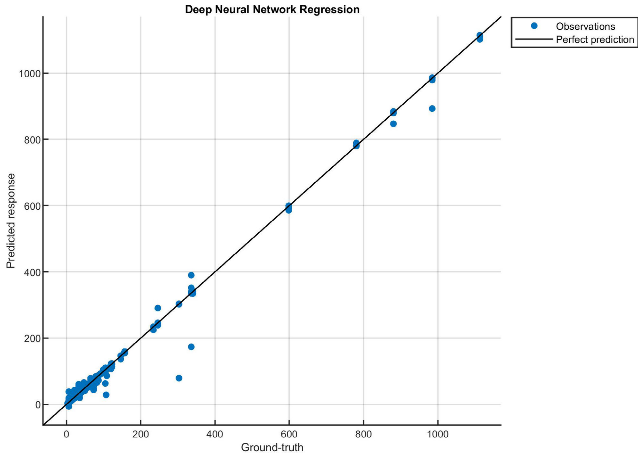


FIGURE 10. Predicted vs ground-truth values plot along with best-fit regression line.

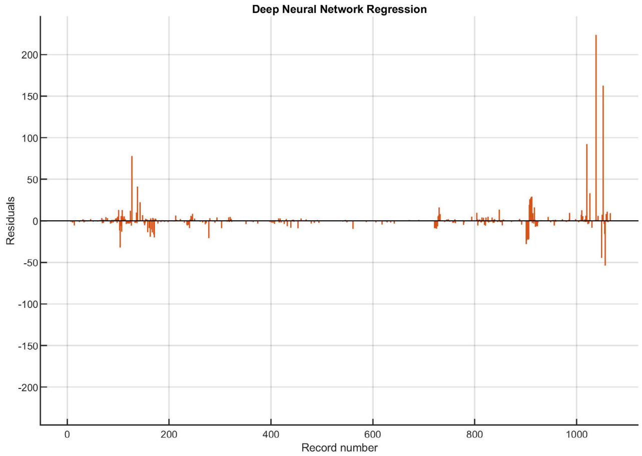


FIGURE 11. Residual plot against observations indicating the amount of error in a typical observation.

In an ideal situation, the plot of the residuals should show a random distribution of values that are relatively uniformly distributed around the identity line. The residual plot offers a practical method for determining the amount of error and the observations that contribute to the error. It can be seen clearly in the plot of Figure 11 that the records number 1000 to 1100 have a larger residual in comparison to the rest of the data. Records in the latter part of the dataset represent the most recent observations, which correlate less with the historical records and, as a result, contain a larger amount of error. The latter records are consequently influenced by other contributing factors such as technology advancements, which are not yet represented by the attributes that are recorded in the dataset.

To provide a closer picture of the predictions made by the deep regression model, Table 6 reports ten randomly drawn observations from the dataset with their ground truth and predicted values. The ground truth is the actual weight of fresh tea leaves collected during a period in a specified area.

TABLE 6. Ten random observations from the data with actual and predicted values of yield.

#	Max Temp	Ave Temp	Min Temp	Ave Humidity	Rainfall	Actual Yield	Predicted Yield
1	35.5	28.4	21.3	28.0	292.5	15.00	15.23
2	34.3	27.1	19.8	31.5	7.0	122.50	122.27
3	30.8	19.9	8.9	20.0	91.0	65.75	65.77
4	34.7	24.2	13.6	34.0	52.0	41.00	41.40
5	33.5	27.9	22.2	30.5	213.8	57.25	57.30
6	33.6	25.0	16.4	31.5	52.0	48.50	48.50
7	33.2	28.5	23.8	42.5	368.8	13.00	13.48
8	38.7	27.4	16.1	45.0	30.0	20.25	19.95
9	28.1	18.5	8.8	20.0	91.0	21.75	21.89
10	33.8	25.2	16.6	34.0	14.0	24.50	24.62

It is to be noted that the NDVI values are not included in the table as they do not convey any useful information to the reader in tabular form. The values of agrometeorological attributes can correlate with the predictions and serve as conclusive evidence. Moreover, the difference (error) in the ground-truth value of yield and predicted value can also be noted.

V. CONCLUSION

Crop yield estimation is widely used to estimate the amount of yield, which is important to ensure food security. There are various approaches in practice to perform this task, and they are based on agrometeorological data, soil characteristics, remote sensing, and crop simulation models. Tea is a perennial crop with different features from seasonal crops, and most yield estimation attempts are made using seasonal crops. On the other hand, this study uses multimodal data, incorporating various data types to perform regression modeling. The agrometeorological data is collected from the tea research institute, which monitors plant health and records the agrometeorological parameters. Moreover, the tea produced on farms is collected and processed separately to ensure accurate measurement of the amount of yield. The data obtained from the tea research institute is supplemented with remote sensing data of Landsat-8 to calculate NDVI to enrich the feature set for improvement of predictive accuracy. The data is processed and prepared for regression modeling, and sixteen regression models are trained and evaluated using 10-fold cross-validation. It was proposed that an optimized neural network can perform tea yield prediction better than conventional regression algorithms due to the data's multimodality and the problem's non-linear nature. The proposed model is constructed through neural architecture search, which defines a search space and performs a search using Bayesian optimization. The optimization approach is intuitive as the grid search is an exhaustive method and requires a lot of computational time. Random search provides a sub-optimal solution; therefore, a limited number of iterations cannot guarantee a good solution. On the other hand, Bayesian optimization has provided a reasonable solution with less time. The model has provided a coefficient of determination (R-squared) of 0.99 with a Mean Square Error (MSE) of 108.17 kg/ha, Root Mean Square Error (RMSE) of 10.87 kg/ha, Mean Absolute Error (MAE) of 2.26 kg/ha and Mean Absolute Percentage Error of 2.92.

The proposed study has explored neural architecture search the lowest prediction error is reported in contrast to various regression modeling approaches. It is concluded that problems of complex and highly non-linear nature can be efficiently modeled using artificial neural networks and performing a neural architecture search can be fruitful in many problems. We also suggest incorporating a larger amount of data both in terms of the number of years and the number of farms so the variabilities introduced to the change of farms or passage of time can be learned and the designed model can generalize better. The model can be further enriched by introducing additional predictors in the data such as attributes about soil fertility and agro management practices as they can have an effect on the amount of yield. Dataset analysis through trend monitoring, feature correlation, and other analysis can be conducted to understand important aspects of the yield such as the predictor's importance. These insights from the data can be helpful in improving the yield prediction, feature selection, and farm-level management of these attributes. As artificial neural networks are observed to be better candidates for modeling complex and highly nonlinear problems and therefore their use is recommended for future studies as well. The introduction of advanced neural network architectures such as long-short term memory networks, gated recurrent units, convolutional neural networks, and attention mechanisms is also expected to be helpful in modeling the problem. The challenges of data scarcity can be handled through the collection of larger amounts of data and performing data augmentation.

APPENDIX A
CLASSIFICATION ALGORITHMS

A. CONVENTIONAL REGRESSION ALGORITHMS

1) MULTIPLE LINEAR REGRESSION (MLR)

Regression models describe the relationship between variables by fitting a line to previously collected data. In contrast to the logistic and nonlinear regression models, the linear regression technique does not include the use of a curved line. The use of regression allows one to estimate the change that occurs in a dependent variable as a result of changes to one or more independent variables. Utilizing a statistical technique known as simple linear regression, one may ascertain the degree of correlation that exists between two numerical variables. To determine the level of correlation that exists between two variables, one technique that may be utilized is known as simple linear regression.

It is necessary to employ multi-linear regression with interactions to carry out multiple linear regression analyses when there are interaction effects present between the various characteristics being investigated. Since there are factors that interact with one another, such as rainfall and humidity, both of which influence yield, it is necessary to do multiple linear regression with the inclusion of interactions term.

Similarly, attribute selection is carried out throughout the process of model development using stepwise linear regression. The multiple linear regression model is trained in this method through an iterative process in which an independent variable is selected for inclusion in the final model. The procedure comprises iteratively adding or removing potential explanatory variables and evaluating their statistical significance after each adjustment. In our implementation, we employed a bi-direction search using forward and backward stepwise regression. This strategy frequently outperforms multiple linear regression, although the computing cost increases dramatically as the number of predictors increases.

2) DECISION TREES (DT)

The decision tree technique is one of the most widely used and successful approaches in the field of supervised learning. It is possible to use it to solve problems involving both regression and classification. It is a predictor that takes the form of a tree and has three distinct categories of nodes inside its structure. The first level at which a sample may be partitioned is called the root node. This level initially represents the whole sample and can be partitioned based on criteria such as entropy and information gain. The decision rules are represented by the branches, while the features of the dataset are represented by the interior nodes of the tree. And last, the regression, which is the ultimate representation of the outcome, is represented by the leaf nodes.

3) SUPPORT VECTOR REGRESSION (SVR)

Support Vector Regression (SVR) is a type of supervised learning algorithm that may be beneficial for generating predictions about numeric data. The same idea that motivates Support Vector Machine (SVM) also drives SVR. The primary goal of the SVR technique is to determine which regression line performs the best fitting. When performing an SVR, the best-fitting line is always the hyperplane that contains the most points. In contrast to other regression models, the SVR's purpose is not to reduce the difference in error between the actual value and the projected value, but rather to identify the best line that can be fitted within a specified threshold. The dividing line is the point when the distance between the hyperplane and the boundary line is minimal. The fit time complexity of SVR grows at a rate that is more than quadratically with the number of samples, significantly limiting its applicability to datasets with more than a few tens of thousands of data points.

Since SVR is a linear regression model, kernel functions are utilized to transform data to describe non-linear relationships. These kernel functions are often used in conjunction with a linear classifier to address non-linear situations. To do linear modeling on non-linear data, we employ kernel functions to project the data into a higher-dimensional space where the features may be linearly separated. This evaluation process incorporates experimenting with several kernel functions, including linear, quadratic, cubic, and Gaussian, to generate yield forecasts.

4) GAUSSIAN PROCESS REGRESSION (GPR)

The Gaussian process regression is a probabilistic supervised learning technique that is becoming more popular for use in many applications involving regression. The performance of a Gaussian processes regression (GPR) model may be evaluated by generating an uncertainty measure, and predictions can be made by employing kernel functions, which provide prior knowledge of a GPR model. The design of the model got inspiration from a wide variety of subfields within mathematics, such as multivariate normal distribution, kernels, a non-parametric model, as well as joint and conditional probability. This approach works particularly well for solving complex regression problems when there is a limited amount of training data. In addition, much as with SVR, the performance of GPR is extremely dependent on kernel functions, which are what enable it to make use of the probability of incorrect predictions. In our implementation, we have tried out a few different kernel functions, including the rational quadratic, squared exponential, Mattern 5/2, and exponential, among others.

B. ENSEMBLE LEARNING

The goal of ensemble learning is to solve a predictive modeling problem by skillfully generating and combining various models. The purpose of ensemble learning is to boost a model's prediction performance or minimize the possibility of selecting a poorly performing model by accident. The most often used methods of ensemble learning are boosting and bagging. In bagging or bootstrap aggregation, several regression algorithms are trained simultaneously on various subsets of the training dataset before being averaged together using either simple or weighted criteria. Boosting, on the other hand, resamples data strategically to provide the most useful training data for each successive predictor, and hence trains on subsampled data.

1) RANDOM FOREST

Random forests are one of the most popularly employed bagging ensemble approaches, and they are used for both the regression and classification methods. The operation is carried out by constructing multiple decision trees based on bootstrapped samples of training data. The output of the random forest can be obtained by taking the average of the values that are provided by the candidate models [26]. Random forests overcome the limitations of decision trees, which frequently suffer from the problem of overfitting on training data [27]. The performance of random forests is better than that of decision trees, but it is frequently inferior to that of gradient-boosted trees [27]. It is important to point out that the performance is heavily dependent on the characteristics of the dataset, and these characteristics cannot be estimated before the model is evaluated.

2) GRADIENT BOOSTING

Gradient boosting is one of the most well-known and frequently used ensemble learning algorithms for regression and classification applications. The term ''gradient boosting'' refers to creating a collectively strong model by combining numerous other weak models. Gradient boosting is an extension of boosting that organizes the construction of weak models through a gradient descent approach over an objective function. The error rates of the model prediction are reduced by gradient boosting, which uses the previously predicted values to lead the parameters of the next model. The gradient of the error in terms of prediction (therefore the name ''gradient boosting'') is used to decide the intended outcome for each instance. The actual performance can only be determined after evaluation as it depends on the peculiarities of the dataset, but in general, gradient boosting outperforms random forest.

3) XGBoost

Extreme Gradient Boosting, or XGBoost, is a distributed gradient-boosted decision tree machine learning toolkit. It is the most widely used machine learning library for classification, regression, and ranking problems, and it differs from traditional gradient-boosted decision trees in that it provides an efficient and parallel tree boosting. XGBoost is a scalable and highly accurate version of gradient boosting that pushes the boundaries of processing power for boosted tree methods. Its primary purpose was to improve the performance and computational speed of machine learning models. In contrast to gradient-boosted decision trees, which produce trees one at a time in sequential order, XGBoost develops its trees in parallel. It uses a level-wise technique, scanning across gradient values and applying partial sums, rather than analyzing the quality of splits at each viable split in the training set, enabling it to evaluate split quality considerably faster.

4) MEAN SQUARED ERROR (MSE) AND ROOT MEAN SQUARED ERROR (RMSE)

Mean squared error is a measure of the average squared difference between the predicted and actual values. The first step in determining the mean's standard error is to square the ''errors'' that separate each data point from the regression line. Any negative values that appeared as a result of finding the difference are eliminated during the squaring step. MSE is almost always positive, either as a result of chance or because the estimator ignores information that may improve precision. In the MSE algorithm, larger mistakes are punished more severely because of the higher attention placed on them. An evaluation of a prediction algorithm's accuracy may be made using an MSE, where a lower MSE implies a more accurate prediction. The formula for the calculation of MSE is given in Eq. 2.

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \bar{y}_i)^2 \tag{2}$$

The Root-Mean-Square Error is the residuals' standard deviation expressed as a square root. The RMSE is a statistic for measuring the variance of residuals, which are numbers that indicate how much the data points deviate from the regression line. In other words, it reveals how tightly the data follows the trend line. In the domains of estimation and regression analysis, the root mean square error is often used as a tool for verifying experimental findings. The formula for the calculation of RMSE is given in Eq. 3.

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{n}(y_i - \bar{y_i})^2}{n}} \tag{3}$$

5) MEAN ABSOLUTE ERROR (MAE) AND MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

When expressing the difference between two measurements, the absolute error is written as a modulus to rule out the possibility of a negative result. The MAE is an alternative to the MSE that uses an absolute value instead of a square root to calculate the error. By doing so, we may prevent the negative effects of squaring the error components. One way to measure the accuracy of a prediction is to take an average of the absolute differences between the projected value and the ground-truth value. The formula for the calculation of MAE is given in Eq. 4.

$$MAE = \frac{\Sigma_{i=1}^{n} \|y_i - \bar{y_i}\|}{n} \tag{4}$$

Mean Absolute Percentage Error on the other hand provides the prediction accuracy of a regression algorithm and is defined by a ratio provided by Eq. 5

$$MAPE = \sum_{i=1}^{n} \left\| \frac{y_i - \bar{y}}{y_i} \right\| \tag{5}$$

6) COEFFICIENT OF DETERMINATION (R^2)

R^2 is a statistical measure that represents the fraction of the variance of a dependent variable, which is explained by one or more independent variables. While the strength of the relationship between an independent variable and a dependent variable is explained by correlation, R^2 explains how one variable's variance explains the second variable's variance. Therefore, about half of the observed variation can be explained by the model, if the R^2 of a model is 0.50. R^2 is calculated using the formula of Eq. 6, where SS_res is the sum of squares of the residuals and SS_total is the total of the squares, which is equal to the variance.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{6}$$

ACKNOWLEDGMENT

The authors would like to thank National Tea and High-Value Crop Research Institute (NTHRI) for their support in providing the data on tea fields.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AVAILABILITY OF DATA AND MATERIALS

The data underlying this article will be shared on reasonable request to the corresponding authors.

CODE AVAILABILITY

The code underlying this article will be shared on reasonable request to the corresponding authors.

REFERENCES

[1] A. A. Hassaballa, A. N. Matori, and H. Z. M. Shafri, "Surface moisture content retrieval from visible/thermal infrared images and field measurements," *Caspian J. Appl. Sci. Res.*, to be published.

[2] D. W. Aha and R. L. Bankert, "A comparative evaluation of sequential feature selection algorithms," in *Proc. Pre 5th Int. Workshop Artif. Intell. Statist.*, 1995, pp. 1–7.

[3] N. Ahmed and H. M. S. Asif, "Ensembling convolutional neural networks for perceptual image quality assessment," in *Proc. 13th Int. Conf. Math., Actuarial Sci., Comput. Sci. Statist. (MACS)*, Dec. 2019, pp. 1–5.

[4] N. Ahmed, H. M. S. Asif, and H. Khalid, "PIQI: Perceptual image quality index based on ensemble of Gaussian process regression," *Multimedia Tools Appl.*, vol. 80, no. 10, pp. 15677–15700, Apr. 2021.

[5] N. Ahmed, H. M. S. Asif, G. Saleem, and M. U. Younus, "Development of crop yield estimation model using soil and environmental parameters," 2021, *arXiv:2102.05755*.

[6] N. Ahmed, H. M. S. Asif, G. Saleem, and M. U. Younus, "Image quality assessment for foliar disease identification (agropath)," *J. Agric. Res.*, vol. 59, no. 2, pp. 177–186, 2021.

[7] A. D. Báez-González, P.-Y. Chen, M. Tiscareño-López, and R. Srinivasan, "Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico," *Crop Sci.*, vol. 42, no. 6, pp. 1943–1949, Nov. 2002.

[8] A. D. Baez-Gonzalez, J. R. Kiniry, S. J. Maas, M. L. Tiscareno, J. Macias C., J. L. Mendoza, C. W. Richardson, J. Salinas G., and J. R. Manjarrez, "Large-area maize yield forecasting using leaf area index based yield model," *Agronomy J.*, vol. 97, no. 2, pp. 418–425, Mar. 2005.

[9] S. K. Bala and A. S. Islam, "Correlation between potato yield and MODIS-derived vegetation indices," *Int. J. Remote Sens.*, vol. 30, no. 10, pp. 2491–2507, May 2009.

[10] D. Batool, M. Shahbaz, H. S. Asif, K. Shaukat, T. M. Alam, I. A. Hameed, Z. Ramzan, A. Waheed, H. Aljuaid, and S. Luo, "A hybrid approach to tea crop yield prediction using simulation models and machine learning," *Plants*, vol. 11, no. 15, p. 1925, Jul. 2022.

[11] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011.

[12] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 801–818.

[13] B. Dhekale, P. Sahu, K. Vishwajith, P. Mishra, and M. Noman, "Modeling and forecasting of tea production in West Bengal," *J. Crop Weed*, vol. 10, no. 2, pp. 94–103, 2014.

[14] P. C. Doraiswamy, S. Moulin, P. W. Cook, and A. Stern, "Crop yield assessment from remote sensing," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 6, pp. 665–674, Jun. 2003.

[15] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 1997–2017, 2019.

[16] M. Feurer and F. Hutter, "Hyperparameter optimization," in *Automated Machine Learning*. Cham, Switzerland: Springer, 2019, pp. 3–33.

[17] C. Funk and M. E. Budde, "Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 115–125, Jan. 2009.

[18] N. Gandhi, O. Petkar, and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks," in *Proc. IEEE Technol. Innov. ICT Agricult. Rural Develop. (TIAR)*, Jul. 2016, pp. 105–110.

[19] N. Gat, H. Erives, G. J. Fitzgerald, S. R. Kaffka, and S. J. Maas, "Estimating sugar beet yield using AVIRIS-derived indices," in *Proc. Summaries 9th JPL Airborne Earth Sci. Workshop*, Pasadena, CA, USA, 2000, pp. 1–10.

[20] K. Gavahi, P. Abbaszadeh, and H. Moradkhani, "DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115511.

[21] S. GopalaPillai and L. Tian, "In-field variability detection and spatial yield modeling for corn using digital aerial imaging," *Trans. ASAE*, vol. 42, no. 6, pp. 1911–1920, 1999.

[22] S. M. E. Groten, "NDVI—crop monitoring and early yield assessment of Burkina Faso," *Int. J. Remote Sens.*, vol. 14, no. 8, pp. 1495–1515, May 1993.

[23] Y. Ul Haq, M. Shahbaz, H. S. Asif, A. Al-Laith, W. Alsabban, and M. H. Aziz, "Identification of soil type in Pakistan using remote sensing and machine learning," *PeerJ Comput. Sci.*, vol. 8, Oct. 2022, Art. no. e1109.

[24] A. A. Hassaballa and A. B. Matori, "The estimation of air temperature from NOAA/AVHRR images and the study of NDVI-Ts impact: Case study: The application of split-window algorithms over (Perak Tengah Manjong) area, Malaysia," in *Proc. IEEE Int. Conf. Space Sci. Commun. (IconSpace)*, Jul. 2011, pp. 20–24.

[25] A. A. Hassaballa, O. F. Althuwaynee, and B. Pradhan, "Extraction of soil moisture from RADARSAT-1 and its role in the formation of the 6 December 2008 landslide at Bukit Antarabangsa, Kuala Lumpur," *Arabian J. Geosci.*, vol. 7, no. 7, pp. 2831–2840, Jul. 2014.

[26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer-Verlag, 2009, doi: 10.1007/978-0-387-84858-7.1998.

[27] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. Cham, Switzerland: Springer, 2009.

[28] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, Aug. 1988.

[29] F. Hutter, B. Kégl, R. Caruana, I. Guyon, H. Larochelle, and E. Viegas, "Automatic machine learning (AutoML)," in *Proc. ICML Workshop Resource-Efficient Mach. Learn., 32nd Int. Conf. Mach. Learn.*, 2015.

[30] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. Int. Conf. Learn. Intell. Optim.* Cham, Switzerland: Springer, 2011, pp. 507–523.

[31] H. Jayanthi, *Airborne and Ground-Based Remote Sensing for the Estimation of Evapotranspiration and Yield of Bean, Potato, and Sugar Beet Crops*. Logan, UT, USAL Utah State Univ., 2004.

[32] M. D. Johnson, W. W. Hsieh, A. J. Cannon, A. Davidson, and F. Bédard, "Crop yield forecasting on the Canadian prairies by remotely sensed vegetation indices and machine learning methods," *Agricult. Forest Meteorol.*, vols. 218–219, pp. 74–84, Mar. 2016.

[33] S. J. J. Jui, A. A. M. Ahmed, A. Bose, N. Raj, E. Sharma, J. Soar, and M. W. I. Chowdhury, "Spatiotemporal hybrid random forest model for tea yield prediction using satellite-derived variables," *Remote Sens.*, vol. 14, no. 3, p. 805, Feb. 2022.

[34] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN framework for crop yield prediction," *Frontiers Plant Sci.*, vol. 10, p. 1750, Jan. 2020.

[35] M. S. Khan, M. Semwal, A. Sharma, and R. K. Verma, "An artificial neural network model for estimating mentha crop biomass yield using Landsat 8 OLI," *Precis. Agricult.*, vol. 21, no. 1, pp. 18–33, Feb. 2020.

[36] P. Kumar, R. Prasad, D. K. Gupta, V. N. Mishra, A. K. Vishwakarma, V. P. Yadav, R. Bala, A. Choudhary, and R. Avtar, "Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data," *Geocarto Int.*, vol. 33, no. 9, pp. 942–956, Sep. 2018.

[37] K. Kuwata and R. Shibasaki, "Estimating crop yields with deep learning and remotely sensed data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 858–861.

[38] D. W. Lamb, M. M. Weedon, and L. J. Rew, "Evaluating the accuracy of mapping weeds in seedling crops using airborne digital imaging: Avena spp. In seedling triticale," *Weed Res.*, vol. 39, no. 6, pp. 481–492, Dec. 1999.

[39] A. Laycock, *Irrigation Systems: Design, Planning and Construction*. Wallingford, U.K.: CABI, 2007.

[40] X. Li, H. Geng, L. Zhang, S. Peng, Q. Xin, J. Huang, X. Li, S. Liu, and Y. Wang, "Improving maize yield prediction at the county level from 2002 to 2015 in China using a novel deep learning approach," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107356.

[41] L. Liang, L. Di, L. Zhang, M. Deng, Z. Qin, S. Zhao, and H. Lin, "Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method," *Remote Sens. Environ.*, vol. 165, pp. 123–134, Aug. 2015.

[42] T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation*. Hoboken, NJ, USA: Wiley, 2015.

[43] W. T. Liu and F. Kogan, "Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices," *Int. J. Remote Sens.*, vol. 23, no. 6, pp. 1161–1179, Jan. 2002.

[44] D. B. Lobell, J. I. Ortiz-Monasterio, G. P. Asner, R. L. Naylor, and W. P. Falcon, "Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape," *Agronomy J.*, vol. 97, no. 1, pp. 241–249, Jan. 2005.

[45] D. B. Lobell, W. Schlenker, and J. Costa-Roberts, "Climate trends and global crop production since 1980," *Science*, vol. 333, no. 6042, pp. 616–620, Jul. 2011.

[46] S. O. Los, *Linkages Between Global Vegetation and Climate: An Analysis Based on NOAA Advanced Very High Resolution Radiometer Data*. Washington, DC, USA: NASA Center for AeroSpace Information, 1998.

[47] A. Magri, H. M. Van Es, M. A. Glos, and W. J. Cox, "Soil test, aerial image and yield data as inputs for site-specific fertility and hybrid management under maize," *Precis. Agricult.*, vol. 6, no. 1, pp. 87–110, Feb. 2005.

[48] F. Arefeen Mila, M. Noorunnahar, A. Nahar, D. C. Acharjee, M. Tania Parvin, and R. J. Culas, "Modelling and forecasting of tea production, consumption and export in Bangladesh," *Current Appl. Sci. Technol.*, vol. 22, no. 2, p. 20, Jul. 2021.

[49] T. Miura, A. R. Huete, and H. Yoshioka, "Evaluation of sensor calibration uncertainties on vegetation indices for MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1399–1409, May 2000.

[50] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104859.

[51] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[52] P. Phan, N. Chen, L. Xu, and Z. Chen, "Using multi-temporal MODIS NDVI data to monitor tea status and forecast yield: A case study at Tanuyen, Laichau, Vietnam," *Remote Sens.*, vol. 12, no. 11, p. 1814, Jun. 2020.

[53] S. Prince, "High temporal frequency remote sensing of primary production using NOAA AVHRR," NOAA (Nat. Ocean. Atmos. Admin.) AVHRR (Adv. Very High Resolution Radiometer), USA, Tech. Rep., 1990, pp. 169–183.

[54] E. E. Raj, K. V. Ramesh, and R. Rajkumar, "Modelling the impact of agrometeorological variables on regional tea yield variability in south Indian tea-growing regions: 1981–2015," *Cogent Food Agricult.*, vol. 5, no. 1, Jan. 2019, Art. no. 1581457.

[55] M. S. Rasmussen, "Operational yield forecast using AVHRR NDVI data: Reduction of environmental and inter-annual variability," *Int. J. Remote Sens.*, vol. 18, no. 5, pp. 1059–1077, Mar. 1997.

[56] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 4780–4789.

[57] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.

[58] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. V. V. Prasad, and I. A. Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in Southern Brazil," *Agricult. Forest Meteorol.*, vol. 284, Apr. 2020, Art. no. 107886.

[59] G. B. Senay, A. D. Ward, J. G. Lyon, N. R. Fausey, and S. E. Nokes, "Manipulation of high spatial resolution aircraft remote sensing data for use in site-specific farming," *Trans. ASAE*, vol. 41, no. 2, pp. 489–495, 1998.

[60] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[61] J. Sun, L. Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model," *Sensors*, vol. 19, no. 20, p. 4363, Oct. 2019.

[62] G. Tan and R. Shibasaki, "Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration," *Ecolog. Model.*, vol. 168, no. 3, pp. 357–370, Oct. 2003.

[63] G. Thomas, J. Taylor, and G. Wood, "Mapping yield potential with remote sensing," *Proc. 1st Eur. Conf. Precis. Agricult., Warwick Univ. Conf. Centre*, Coventry, U.K., 1997.

[64] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 847–855.

[65] J. Vanschoren, "Meta-learning," in *Automated Machine Learning*. Cham, Switzerland: Springer, 2019, pp. 35–61.

[66] M. Wistuba, A. Rawat, and T. Pedapati, "A survey on neural architecture search," 2019, *arXiv:1905.01392*.

[67] C. Yang and G. L. Anderson, "Mapping grain sorghum yield variability using airborne digital videography," *Precis. Agricult.*, vol. 2, no. 1, pp. 7–23, 2000.

[68] P. Yang, G. Tan, Y. Zha, and R. Shibasaki, "Integrating remotely sensed data with an ecosystem model to estimate crop yield in North China," in *Proc. 20th ISPRS Congr. Commission VII, WG VII/2*, 2004, pp. 150–156.

[69] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," 2016, *arXiv:1611.01578*.

[70] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.

ZEESHAN RAMZAN is currently pursuing the Ph.D. degree in computer science with a focus on machine learning. He has been teaching, since 2016. He is also a Lecturer in computer science with the University of Engineering & Technology (New Campus) Lahore, Pakistan. He is an active academic researcher with research interests, including e-agriculture, open-source machine-learning and deep-learning-based application development, and health Informatics.

H. M. SHAHZAD ASIF received the B.S. and M.S. degrees in computer science from the University of Engineering & Technology Lahore, in 2003 and 2007, respectively, and the Ph.D. in informatics from the University of Edinburgh, U.K., in 2012. He has been a Faculty Member with the Department of Computer Science and Engineering, University of Engineering & Technology Lahore, since 2004. He has over five years of experience in web and desktop application development. His research interests include utilizing machine learning and allied disciplines for solving real-world problems found in structured or unstructured data.

IRFAN YOUSUF received the Ph.D. degree from the University of Science & Technology, Daejeon, South Korea. He is currently an Assistant Professor with the Department of Computer Science (New Campus), University of Engineering & Technology Lahore (UET Lahore). He was a Postdoctoral Fellow with the Korea Institute of Science & Technology, Seoul, South Korea, before joining UET Lahore. His research interests include graph mining, network analysis, graph compression, and peer-to-peer systems.

MUHAMMAD SHAHBAZ received the Ph.D. degree from Loughborough University, U.K. He is currently a Full Professor with the Department of Computer Science and Engineering, University of Engineering & Technology Lahore. He has delivered several talks in the industry at national and international levels and various conferences around the world. He has a wide experience in the field of data science and has published more than 60 articles in the same domain. His research interests include healthcare informatics, fog computing, data science, and artificial intelligence.