Potato Research (2025) 68:4327–4355
https://doi.org/10.1007/s11540-025-09924-3

Investigating Fine-tuning Strategies in Statistical and Machine Learning Models for Developing Region-Specific Potato Yield Prediction Models

Ajith S[1] · Manoj Kanti Debnath[1] · Deb Sankar Gupta[1] · Deepranjan Sarkar[2] · Subhendu Bandyopadhyay[3] · Partha Sarathi Patra[4] · Rakesh S[5]

Received: 20 March 2025 / Accepted: 13 August 2025 / Published online: 12 September 2025

Abstract

Accurate crop yield prediction, especially for potatoes with high price volatility, is essential for informed market decisions, maintaining supply–demand balance, preventing distressed sales, and stabilizing prices. As weather impacts crop yield more significantly at the regional level, district-level modelling is ideal. Accurate predictions require proper model selection, data preparation, feature selection and parameter tuning. This study introduced path coefficient-based weather indices alongside correlation-based indices that weigh weekly weather based on its true effect on potato yield. Three penalized regression models (Ridge regression, *LASSO* and Elastic Net) and two machine learning (ML) models (artificial neural networks [ANN] and support vector regression [SVR]) were evaluated. To prevent overfitting, stepwise regression, principal component analysis, and partial least squares regression (PLSR) were employed for feature selection. The ANN model was tested under three different nonlinear activation functions with optimized learning rates and hidden layer neurons, while the SVR model was tested with three different kernel functions and optimized hyperparameters. The proposed path coefficient-based indices improved yield prediction alongside correlation-based indices. ANN and SVR models with PLSR-selected features outperformed others with the lowest prediction errors. The best-performing configurations were ANN with tangent hyperbolic activation function, learning rate below 0.1 and fewer hidden-layer neurons than input layers and SVR with radial basis function kernel and optimized hyperparameters. This systematic and comprehensive approach, which combines the construction of information-rich weather indices-based input, PLSR-based feature selection and fine-tuning of ML models, effectively captures complex nonlinear weather-crop interactions and is crucial for developing robust location-specific potato yield prediction models.

Extended author information available on the last page of the article

Keywords Activation function · Artificial neural network · Partial least squares regression · Penalized regression models · Potato yield prediction · Support vector regression

Introduction

Accurate crop yield prediction is crucial for stakeholders across all levels of agriculture, from individual farmers to national and international organizations, aiding in policy-making, resource management and market decisions (Lobell et al. 2006; Vander-Velde et al. 2019). Governments rely on yield forecasts for food security, crop insurance, input supply, and trade policies (Kross et al. 2020; Li et al. 2022). Farmers use predictions for crop selection, acreage allocation, other farm management decisions, and economic planning (Basha et al. 2020; Ansarifar et al. 2021). Yield prediction, especially for crops with high price volatility like potatoes, is crucial for making informed market decisions, maintaining supply–demand balance, preventing distressed sales, and stabilizing prices (Azizabadi and Badreldin 2025; Jégo et al. 2025). Localized forecasting of potato yields helps farmers make informed decisions about the timing and location of their product sales and planning storage thereby maximizing their profits (Lin et al. 2023; Mukiibi et al. 2025; Tamayo-Vera et al. 2025). The regional forecasts support location-specific decision-making (Van Klompenburg et al. 2020; Rai et al. 2022).

Climate change has a significant impact on global, regional and local crop yields, with more pronounced effects at smaller scales (Trnka et al. 2016; Hao et al. 2018; Jia et al. 2022). While large-scale production declines may go unnoticed, regional agricultural losses pose serious concerns (Bhardwaj et al. 2022). The impact of climate change on potato yield is likely to vary based on regional conditions, necessitating regional-level strategies (Bender and Sentelhas 2020). Crop growth is closely linked to weather conditions in which the crop is grown (Haverkort and Verhagen 2008). High temperatures promote vegetative growth, delay tuberization, and result in smaller tubers with lower yields (Haverkort 1987; Singh et al. 2013). Interannual variations in weather conditions lead to differences in potato yields (Li and Zhang 2020). Integrating meteorological data into potato yield prediction models is essential. Weather-based models offer reliable crop yield forecasts, with various climate-crop models assessing the impact of climate change on agriculture (Rötter et al. 2011).

Since crop yield is influenced by complex environmental interactions, a model effective in one location or crop may not perform well elsewhere (Ajith et al. 2025). Given the complexity of the effect of weather on crop yield, exploring different functional models is essential for optimal location-specific predictions (Mavromatis 2016). Statistical models help analyze complex relationships between agricultural systems and climate variables (Roberts et al. 2017; Siebert et al. 2017). Machine learning (ML) enhances crop yield prediction by identifying patterns in climate response (Kang et al. 2020). ML models optimize regional yield predictions and adapt to climate change effects (Paudel et al. 2021; Hussain et al. 2021), offering better performance with limited data (Batool et al. 2022). Artificial neural network

(ANN) effectively model nonlinear relationships, influencing performance by learning algorithms, activation functions, learning rate and hidden layer neurons (Vora and Yagnik 2014; Hunter et al. 2012). Support vector machine (SVM), initially designed for classification (Vapnik 2000), evolved into support vector regression (SVR) for predictive modeling. The performance of SVR depends on selecting an optimal kernel function and tuning hyperparameters for better yield prediction using weather data (Qu and Zhang 2016; Paidipati et al. 2021). The availability of optimization options in ANN and SVR models enables fine-tuning models to capture location-specific patterns in the association between weather conditions and crop yield (Ajith et al. 2024).

Both technological advancements and weather variability influence crop yield. While high-yielding varieties and improved practices contribute to a steady yield increase, weather fluctuations within and between seasons create unpredictable yield variations (Agrawal and Mehta 2007). Weather conditions affect crops differently at various growth stages. For instance, potato tuber induction, initiation, and setting require different temperature ranges (Struik 2007). Hence, assigning appropriate weightage to weekly weather variables improves model accuracy (Jain et al. 1980). Composite weighted weather indices, derived using correlation coefficients, effectively reduce bias in climate-based models and enhance the accuracy of yield predictions (Gupta et al. 2018; Singh et al. 2021).

Weather variables are often correlated with each other and the correlation between a weather factor and yield may fail to provide the exact effect of a given weather factor on crop yield. Path analysis separates direct and indirect effects, providing a clearer understanding of the variable influence on yield (Niles 1923). Unlike correlation coefficients, path coefficients account for inter-variable relationships, making them more precise for weighting weather indices (Hope 1971). Hence, this study proposes a path coefficient-based weighted index to improve accuracy in combination with correlation-based indices.

Variable selection is crucial in predictive modeling, eliminating redundant features to make the algorithm work fast, improve efficiency, reduce overfitting and enhance interpretability (Kumar et al. 2019). This can be achieved through subset selection or transforming variables into a smaller set that retains most of the information in the data. Stepwise regression helps identify statistically significant predictors, but it may overlook causal variables or retain spurious ones, potentially leading to poor generalization of the model. Sometimes, stepwise regression eliminates some variables that contain relevant information on the response variable; complete exclusion of such variables diminishes the overall predictive accuracy of the model (Olusegun et al. 2015). Multicollinearity further complicates model stability (Vatcheva et al. 2016; Daoud 2017).

Principal component analysis (PCA) is a widely used dimensionality reduction technique that generates uncorrelated components while retaining maximum variance (Jolliffe et al. 2003). Hotelling (1957) introduced the idea of replacing original explanatory variables with principal components in statistical models. However, PCA does not account for the response variable, which may lead to the omission of relevant information. Components that explain lower variance in the independent variable set are typically excluded from analysis, although they may still

be important for predicting the response variable (Kung and Sharif 1980; Jolliffe 1982). In contrast, partial least squares regression (PLSR) derives components that maximize covariance between independent and response variables (Firinguetti et al. 2017), making PLSR more effective in capturing key relationships for predictive modeling (Hu et al. 2018).

When developing region-level predictive models for crops such as potato, a key research gap exists in the absence of a systematic approach, starting from construction of predictive and information rich input feature such as composite weather indices from weekly weather data that optimally weigh the effect of weekly weather condition on yield, best dimension reduction technique and fine-tuning strategies for ML models to capture maximum possible information concerning the association between weather and yield. Such a comprehensive and structured approach is essential to ensure robust model generalization and the development of an optimized predictive model tailored to the target location.

Based on the above, the specific objectives of this study were (i) to develop correlation- and path coefficient-based weighted composite weather indices that effectively condense weekly weather data and capture critical weather-yield relationships for potato at the regional scale; (ii) to evaluate suitable dimensionality reduction techniques, such as stepwise regression, PCA and PLSR, to reduce input complexity while minimizing model overfitting; (iii) to fine-tune the key hyperparameters of models, such as ANN, SVR, Ridge regression, least absolute shrinkage and selection operator (*LASSO*) and Elastic Net (*ENET*), to capture complex relationships between weather and potato yield; and (iv) to identify the best-performing model by comparing the performance of the fitted models to determine its optimal hyperparameter ranges and select the most suitable activation or kernel functions. This comprehensive approach was intended to serve as a systematic framework for developing robust, region-specific potato yield prediction models with enhanced accuracy.

Materials and Methods

The study focused on four northern districts of West Bengal, India, located in the eastern sub-Himalayan region, viz. Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur. These four districts produced 3.32 million tons of potatoes every year, from approximately one lakh cultivational areas with an average yield of 31.32 tons ha^{-1}, making this region one of the hotspots for potato production. Yearly potato yield data for these districts from 1997–1998 to 2022–2023 were accessed from the official website of the Directorate of Economics and Statistics, Ministry of Agriculture & Farmers Welfare, Government of India (https://data.desagri.gov.in/website/crops-apy-report-web). Weekly weather data for the same districts were sourced from the Regional Meteorological Centre, Kolkata. The study considered five weather parameters: maximum temperature, minimum temperature, relative Humidity, rainfall and wind speed. Weekly weather data from the 47th standard meteorological week (SMW) of a year to the 11th SMW of the following year, covering the potato growing period were used to develop yearly weather indices.

Calculation of Different Weather Indices

Three types of yearly weather indices were developed. One unweighted index and two weighted indices.

Unweighted Indices

Unweighted indices are the simple average of weekly weather variables of the weeks in which the potato crop is grown.

Correlation Coefficient Based Weighted Weather Indices

The correlation coefficient-based weighted indices are calculated as follows,

$$C_{ij} = \frac{\sum_{k=1}^{m} r_{jk}.X_{ijk}}{\sum_{k=1}^{m} r_{jk}} \quad (1)$$

where C_{ij} is weighted weather index for jth weather parameter in ith year with correlation coefficient as weight, r_{jk} is the correlation coefficient between potato yield and jth weather parameter at kth week and X_{ijk} is the observation of jth weather parameter in kth week of ith year.

Before calculating correlation coefficients, yield data was tested for trend effects. The Modified Mann–Kendall test is a more robust test to detect the presence of a trend with the null hypothesis states that no trend exists (Hamed and Rao 1998). Additionally, Sen's slope, a nonparametric method, was used to estimate trend magnitude by measuring change per unit time (Kuriqi et al. 2020).

If a significant trend was detected, the detrended yield is to be used to calculate correlation coefficients to account for the true effect of individual weather factors on yield (Agrawal et al. 1986). Detrending isolates the actual effect of weather by removing the influence of technological advancements such as improved crop varieties and agronomic practices (Huzsvai et al. 2022). Simple linear regression is commonly used for detrending, where yield is regressed on the year variable. The detrended yield is obtained by subtracting the predicted yield from the actual yield (Sridhara et al. 2020).

Path Coefficient-based Weighted Weather Indices

Path analysis partitions the correlation coefficient into two components, viz., direct and indirect effects (Li 1956; Freedman 1987). The path coefficient accounts for only the direct effect of an explanatory variable on the ultimate response variable (Alwin and Hauser 1975). Hence, a novel weighted index was proposed as a weighted average of weather variables, where the weight is the path coefficient.

$$P_{ij} = \frac{\sum_{k=1}^{m} p_{jk}.X_{ijk}}{\sum_{k=1}^{m} p_{jk}} \quad (2)$$

where P_{ij} is weighted weather index for j^{th} weather parameter in i^{th} year with path coefficient being the weight, and p_{jk} is the path coefficient between the detrended potato yield and j^{th} weather parameter at k^{th} week.

Variable Selection or Dimension Reduction

Since there are five weather parameters, five unweighted indices, five correlation coefficient-based weighted indices and five path coefficient-based weighted indices are developed, as shown in Table 1. Including all explanatory variables in the model increases complexity and requires estimating numerous parameters. Moreover, weather variables are often correlated, leading to multicollinearity, which results in unstable coefficient estimates.

To reduce dimensionality and retain only a few uncorrelated variables that explain the maximum potato yield variability, three variable selection techniques were employed: stepwise regression, PCA and PLSR.

Stepwise Regression

Stepwise regression is a classical variable selection methodology that is used to identify and select a useful subset of the important explanatory variables (Lewis-Beck 1978). Stepwise regression selects explanatory variables based on their statistical significance. In each step, every candidate variable that is already included will be revisited to check whether the inclusion of the variable in the model is still worth it (Pope and Webster 1972). It is possible to drop a variable that is found

Table 1 Description of weighted and unweighted indices

S no	Indices	Description
1	Tmax	Unweighted indices of maximum temperature
2	Tmin	Unweighted indices of minimum temperature
3	RH	Unweighted indices of relative humidity
4	WS	Unweighted indices of windspeed
5	RF	Unweighted indices of rainfall
6	CC_Tmax	Correlation coefficient-based weighted indices of maximum temperature
7	CC_Tmin	Correlation coefficient-based weighted indices of minimum temperature
8	CC_RH	Correlation coefficient-based weighted indices of relative humidity
9	CC_WS	Correlation coefficient-based weighted indices of windspeed
10	CC_RF	Correlation coefficient-based weighted indices of rainfall
11	PC_Tmax	Path coefficient-based weighted indices of maximum temperature
12	PC_Tmin	Path coefficient-based weighted indices of minimum temperature
13	PC_RH	Path coefficient-based weighted indices of relative humidity
14	PC_WS	Path coefficient-based weighted indices of windspeed
15	PC_RF	Path coefficient-based weighted indices of rainfall

to be nonsignificant in the later stages, given that the significant variable is added (Henderson and Denison 1989). The best subset model can be selected using statistical criteria such as the highest Adjusted R^2 or lowest Akaike Information Criteria (*AIC*) values (Ruengvirayudh and Brooks 2016).

Principal Component Analysis

PCA is a multivariate data reduction technique to transform the original set of variables into a smaller set of linear combinations. Most of the variation in the original data set is explained by that linear combination. These linear components are called as principal components (PC).

PCA was carried out using all the developed indices. The decision on the number of PCs to be retained was taken based on criteria, such as proportion of variance explained and Kiser criteria (Kaiser 1960). Let PC_1, PC_2, …, PC_k be first k ($k < 15$) principal components that explain about 80% of the total variation in the original data, and their eigenvalues are greater than one. Then, these k principal components will be used as input variables in the models. As the developed principal components are orthogonal or uncorrelated to each other, the PCA can be efficiently employed to avoid the multicollinearity problem.

Partial Least Square Regression

PCA decomposes the explanatory dataset (X) to obtain components that best explain X rather than the response variable (Y) (Bhat and Vidya 2018). Partial least square regression (PLSR) finds components from X that best predict Y (Hoskuldsson 1998). The PLSR components are obtained by maximizing the covariance between the dependent and explanatory variables in the data set (Abdi 2010). That is

$$\text{Max. Cov}\left(\widetilde{X} \cdot u_h, \overline{Y} \cdot v_h\right) \tag{3}$$

In PLSR, new variables obtained are components or latent variables, which are orthogonal linear combinations of the explanatory variables that are arranged in the order of variation in the response variable they explain (Krishnan et al. 2011). PLSR can be used as an alternative to PCA in the context of dimensionality reduction. The difference between PLSR and PCA is that the response variable is considered while extracting components in the former and the response variable is not considered in the latter case (Jia et al. 2014). A cross-validation procedure is commonly employed in PLSR to determine the appropriate number of components. The optimal number of PLSR components corresponded to a minimal root mean square error for cross-validation (*RMSE-CV*) (Król 2017).

Machine Learning Models

Artificial Neural Network

ANN is a most popular machine learning technique that mimics the mechanism of learning in biological organisms (Kittichotsatsawat et al. 2022). The relationship between the output and the inputs can be mathematically represented as follows:

$$Y_t = f\left\{ \sum_{j=0}^{q} \omega_j \cdot g\left(\sum_{i=1}^{p} \omega_{ij} \cdot x_i \right) \right\} + e_t \tag{4}$$

where p is the number of input nodes; q is the number of hidden nodes; ω_j ($j = 1,2,\ldots q$) and ω_{ij} ($i = 1,2,\ldots p$) are the connection weights; g and f denote the activation function at the hidden and output layers, respectively, and e_t is the error term.

The multilayer perceptron-based ANN captures nonlinear patterns more effectively than linear regression models. ANN performance depends on the learning algorithm, learning rate, activation function in hidden layers and the number of hidden neurons. The backpropagation algorithm used in this study ensures fast convergence while avoiding local minima. It trains the model by adjusting random initial weights to minimize error using gradient descent. Data normalization prevents bias from variable magnitude (Puig-Arnavat and Bruno 2015). In this study, the ANN model is fitted under the following configurations.

1. *Input layer*: number of input neurons equal to SR-selected variables for SR_ANN, PCs for PCA-ANN and PLSR components for PLSR_ANN models.
2. Hidden-layer: The number of neurons in the hidden-layer was optimized by a tenfold cross-validation procedure.
3. Output layer: single neuron (yield prediction).
4. Training algorithm: Backpropagation was used with optimized learning rates, ranging from 0.01 to 1.00 in 0.01 intervals (100 rates). The optimal learning rate (η) was selected, where the gradient descent algorithm minimized the error function.
5. Activation functions tested:

$$\text{Logistic} : f(x) = \frac{1}{(1+e^{-x})} \tag{5}$$

$$\text{Tangent hyperbolic (Tanh)} : f(x) = \frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})} \tag{6}$$

$$\text{Restricted linear unit (ReLU)} : f(x) = \max(0,x) = \begin{cases} x_i \; if \; x_i \geq 0 \\ 0 \; if \; x_i < 0 \end{cases} \tag{7}$$

Evaluating models across three input conditions, three activation functions and optimized learning rates helps tailor ANN models to each location, maximizing the association between weather factors and potato yield.

Support Vector Regression

SVM is a powerful and efficient nonlinear multivariate machine learning technique that maps input feature vectors into a high-dimensional space to construct the optimal separating hyperplane (Nti et al. 2021). SVR extends SVM for regression tasks by incorporating an ϵ-insensitive tube, which helps balance model complexity and prediction error. Minimizing ϵ-loss function flattens the ϵ-insensitive tube, allowing it to encompass most of the training data points. Let y be the dependent variable and x be a set of input data; the support vector regression is given as

$$y = \sum_{i=1}^{n} x_i \cdot \mathrm{w_i} \cdot K\left(x_i, x_k\right) + b + \varepsilon \tag{8}$$

where $K(x_i, x_k)$ is a kernel function. The kernel is a function that simulates the input data into higher-dimension feature space. A linear kernel function can be used when the data is linearly separated. Nonlinear kernels such as radial basis function (*RBF*) and polynomial kernels were appropriate when the data is not linearly separated (Al Azies et al. 2019). The kernel functions tested are

$$\text{Linear kernel}: \; k\left(x_i, x_k\right) = x_i^T \cdot x_k \tag{9}$$

$$\text{Polynomial kernel}: \; k\left(x_i, x_k\right) = \left(\gamma \cdot x_i^T \cdot x_k + 1\right)^d \tag{10}$$

$$\text{RBF kernel}: \; k\left(x_i, x_k\right) = \exp\left(-\gamma \,\|\, x_i - x_k \,\|^2\right) \tag{11}$$

where x_i and x_k are the support vectors. Hyperparameters in SVR are values that govern the learning process. For an SVR model using a linear kernel, the key hyperparameters to optimize are cost (C) and epsilon (ε). In the case of the nonlinear RBF kernel, an additional hyperparameter, gamma (γ), must also be tuned. Similarly, for the polynomial kernel, the degree of the polynomial (d) is another crucial parameter requiring optimization. The parameter C introduces a penalty term that balances the trade-off between the width of the hyperplane margin and the training error. ε defines the margin of the hyperplane, while γ serves as a scaling parameter (Savas and Dovis 2019). A polynomial kernel with degree = 1 functions similarly to a linear kernel, whereas higher-degree polynomials are necessary to model complex nonlinear relationships between features (Hasan et al. 2016).

Optimizing these hyperparameters to minimize the ϵ-loss function is essential for fine-tuning the model to accurately capture the relationship between the response variable and input features. In this study, a grid search with tenfold cross-validation was employed to determine the optimal combination of hyperparameters.

Penalized Regression Models

Linear regression uses ordinary least squares (*OLS*) for parameter estimation. However, multicollinearity destabilizes regression coefficients, increasing

standard errors. Severe multicollinearity complicates estimation and interpretation, while *OLS* may overfit when correlated variables introduce excessive noise. Penalized or regularized regression models can address this through penalization. The general form of the function to be minimized is

$$\sum_{i=1}^{n}\left(y_i-\sum x_{ij}\beta_j\right)^2+\lambda\times Penalty(\beta) \tag{12}$$

The regression coefficients (β) corresponding to collinear variables are also minimized due to the imposition of a penalty term λ (regularization parameter) in the second part of the equation. The nature of this penalty term varies across different penalized regression models. A smaller λ applies a minimal penalty. As $\lambda\rightarrow\infty$, coefficients shrink toward zero. As a result, some coefficients shrink, reducing the overall slope of the fitted line (Wang et al. 2022). This regularization enhances the performance of linear regression models.

1. Ridge regression

Ridge regression imposes L_2 penalty, which is the sum of squared coefficients as given below,

$$S(\beta)_{RR}=\sum_{i=1}^{n}\left(y_{\mathrm{i}}-\sum x_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{p}\left(\beta_j\right)^2 \tag{13}$$

2. Least absolute shrinkage and selection operator

LASSO applies an L_1 penalty by sum of their absolute penalty.

$$S(\beta)_{LASSO}=\sum_{i=1}^{n}\left(y_i-\sum x_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{k}\left|\beta_j\right| \tag{14}$$

The *LASSO* estimation sets some coefficients to zero, retaining only those that minimize the RSS, thus achieving variable selection.

3. Elastic Net

ENET, introduced by Zou and Hastie (2005), is a combination of Ridge and *LASSO* regression in which both sum and absolute penalty are imposed together.

$$S(\beta)_{ENET}=\sum_{i=1}^{N}\left(y_i-\sum x_{ij}\beta_j\right)^2+\lambda\left((1-\alpha)\sum_{j=1}^{p}\left(\beta_j\right)^2+\alpha\sum_{j=1}^{p}\left|\beta_j\right|\right) \tag{15}$$

where $\lambda>0$ and $0<\alpha<1$. The parameter λ controls the balance between the L_1 and L_2 penalties, making *ENET* a generalization of Ridge and *LASSO* regression. If α is close to zero, *ENET* resembles Ridge, whereas values near one make it more like *LASSO*. At $\alpha=0.5$, *ENET* applies an equal mix of both penalties.

The optimal penalty parameter value was determined through cross-validation. Variables with zero coefficients are excluded, while those with non-zero coefficients are retained in the fitted model, effectively achieving variable selection.

Model Evaluation Criteria

Eighty percent of the data was randomly selected for model training, while the remaining 20% was used for validation. Random splitting was employed to ensure that both recent and past years were represented in both training and testing datasets, as the recent years generally exhibit higher yield values. This approach promotes balanced and unbiased model evaluation (Ajith et al. 2023a). ANN and SVR were fitted using each of the three-dimensional reduction techniques, namely SR, PCA and PLSR. Penalized regression models inherently perform variable selection by shrinking coefficients and mitigating multicollinearity, eliminating the need for separate variable selection or dimension reduction. Hence, nine models were fitted for each district.

The models were fine-tuned by optimizing the respective hyperparameters. Fine-tuning the hyperparameters of ML models has become a key strategy for utilizing small datasets by capturing as much information as possible from the available data (Wojciuk et al. 2024). Then validation of the trained model for the unseen data (testing data) evaluated the generalized performance of the models. The best-fitted model for each district was selected based on the model evaluation criteria R^2, mean absolute error (*MAE*), root mean square error (*RMSE*) and normalized root mean square error (*nRMSE*). The model that performed best in both training and validation was considered the best-fit model. The systematic procedure followed in this study is illustrated in Fig. 1.

Results and Discussion

Trend Analysis

The presence of temporal trend in the potato yield was tested using the modified Mann–Kendall test, which accounts for autocorrelation in the time series data and is more robust than the Mann–Kendall test. Further, the change per year in the yield was quantified using Sen's slope method (Table 2). The results revealed that a highly significant and positive trend in potato yield was observed across all districts ($p < 0.01$), as indicated by the modified Mann–Kendall test. As quantified by the Sen's slope estimator, the potato yield increases by 1 ton/year in Malda and over half a ton per year in the other three districts, indicating consistent yield improvement over the study period.

Calculation of Weather Indices

Unweighted indices were computed as the simple average of each weather variable during the growing period. Since a significant increasing trend was detected in yield, detrending was performed prior to the calculation of correlation and path

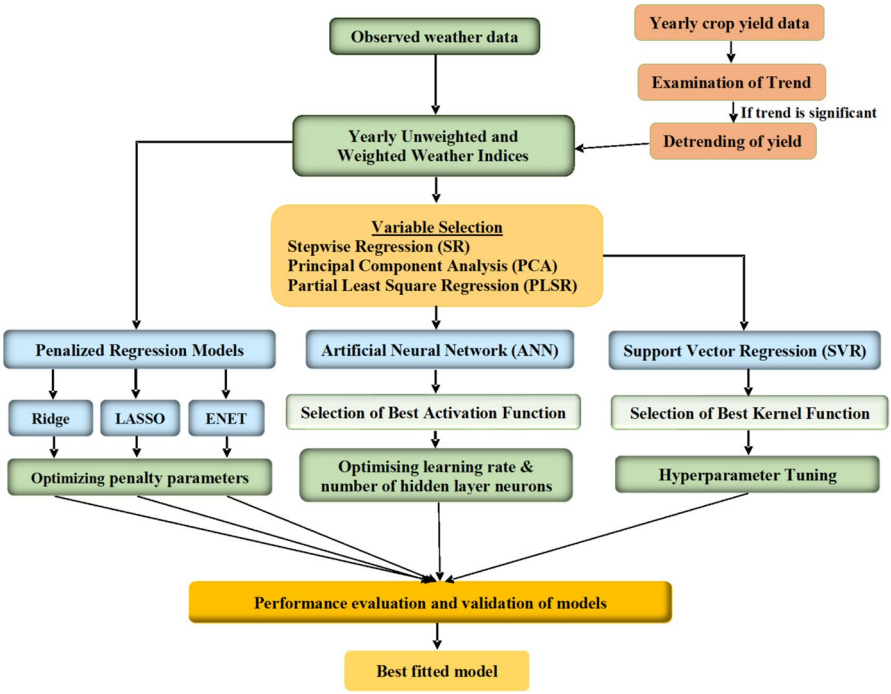

Fig. 1 Systematic framework for developing a robust, region-specific potato yield predictive model

Table 2 The trend in potato yield

District	Average yield (t ha^{-1})	Modified Mann–Kendall test		Sen's slope
		Z statistic	p value	
Cooch Behar	24.91	3.66	< 0.00	0.56
Jalpaiguri	25.83	4.67	< 0.00	0.66
Malda	27.52	3.61	< 0.00	0.99
Uttar Dinajpur	24.45	4.06	< 0.00	0.55

coefficient-based indices to eliminate the influence of time-related changes and store only the true relationship between yield and weekly weather variables in the derived indices. Weighted indices were then derived using formulas (1) and (2).

Dimension Reduction

Stepwise Regression

Stepwise regression was performed using the developed weather indices and a time trend variable as independent variables, with potato yield as the dependent

variable. The most significant input features across the four districts are summarized in Fig. 2.

1. *Cooch Behar*: CC_Tmin was included first (5% level of significance (*LoS*), Adjusted $R^2 = 0.32$, $AIC = 150$). Adding CC_WS increased the adjusted R^2 to 0.56 and lowered *AIC*. CC_RF and the time trend variable were included next, further improving the model (adjusted $R^2 = 0.66$). All four indices were significant at 5% *LoS*.
2. *Jalpaiguri*: Five indices (CC_Tmin, Year, PC_RH, PC_Tmax and PC_WS) were added sequentially, improving the model from an adjusted R^2 of 0.62 ($AIC = 130$) to an adjusted R^2 of = 0.92 with lower *AIC*. All indices and the intercept were significant at 5% *LoS*.
3. *Malda*: Five indices (Time trend, CC_Tmin, PC_RH, RF and PC_Tmax) were included stepwise, leading to adjusted $R^2 = 0.86$ and a final *AIC* of 130. All indices and the intercept were significant at 5% *LoS*.
4. *Uttar Dinajpur*: Four indices (Time trend, CC_Tmin, CC_RF and CC_RH) were included, achieving adjusted $R^2 = 0.88$ with a low *AIC*. All indices were significant at 5% *LoS*.

Path coefficient-based indices, along with correlation-based indices, were found to have a significant effect on the yield in Jalpaiguri and Malda districts. In most

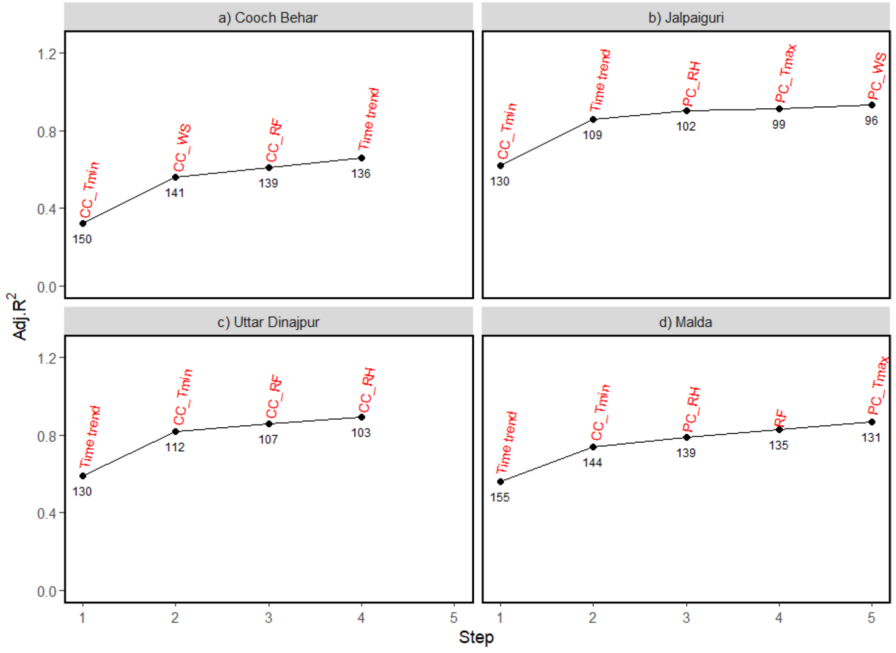

Fig. 2 Summary of stepwise regression results (Values indicated at each step are their corresponding *AIC* values.)

cases, weighted indices were found to have a significant effect on the yield, emphasizing the varying effect of weather factors over the potato growth period.

Principal Component Analysis

PCA was performed for each district using 15 standardized weather indices. The summary is presented in Table 3. PC_1 explained the highest variability, accounting for 29.69% (Cooch Behar), 29.00% (Jalpaiguri), 35.10% (Malda) and 39.43% (Uttar Dinajpur). With the addition of further components, the cumulative variance explained increased substantially. The optimal number of components for model input was determined based on the combined criteria of eigenvalues greater than one and cumulative variance exceeding 80%. Cooch Behar, Jalpaiguri and Uttar Dinajpur had five principal components with eigenvalues greater than one, explaining 83.59%, 83.09% and 86.10% of the total variability, respectively. Malda had four principal components with eigenvalues greater than one, explaining 84.56% of the variability. The components beyond these thresholds contributed less than 6% variability; therefore, only the first five components were selected for further analysis in Cooch Behar, Jalpaiguri, Uttar Dinajpur and the first four components in Malda.

Partial Least Square Regression

PLSR was applied to each district, with potato yield as the dependent variable and 15 weather indices and time trend as independent variables. The optimal number of components was selected using cross-validation, based on the lowest *RMSE-CV*. PLSR results are summarized in Table 4, and cross-validation plots are shown in Supplementary file 2.

1. *Cooch Behar*: *RMSE-CV* = 5.19 at the first component, explained 24.68% of independent variable variability and 66.48% of yield variability. The cross-validation

Table 3 Summary of principal component analysis (PCA)

District	Parameters	PC_1	PC_2	PC_3	PC_4	PC_5
Cooch Behar	Eigen value	4.75	3.00	2.49	1.98	1.15
	Variability Explained (%)	29.69	18.73	15.59	12.38	7.20
	Cumulative Variability (%)	29.69	48.41	64.01	76.39	83.59
Jalpaiguri	Eigen value	4.64	3.20	2.36	1.93	1.17
	Variability Explained (%)	29.00	19.99	14.73	12.03	7.34
	Cumulative Variability (%)	29.00	48.99	63.72	75.75	83.09
Malda	Eigen value	5.62	3.87	2.91	1.13	0.87
	Variability Explained (%)	35.10	24.18	18.20	7.07	5.44
	Cumulative Variability (%)	35.10	59.29	77.49	84.56	90.01
Uttar Dinajpur	Eigen value	4.71	4.08	2.02	1.59	1.37
	Variability Explained (%)	29.43	25.53	12.65	9.91	8.57
	Cumulative Variability (%)	29.43	54.96	67.62	77.53	86.10

Table 4 Summary of partial least square regression (PLSR) analysis

District	Cumulative variability explained	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Cooch Behar	Independent variables (%)	24.68	46.50	56.06	63.80	67.72
	Yield (%)	66.48	73.10	76.79	78.92	80.33
Jalpaiguri	Independent variables (%)	28.50	39.98	47.69	58.46	68.62
	Yield (%)	81.16	89.57	93.19	94.30	95.11
Malda	Independent variables (%)	34.54	46.63	61.67	75.84	86.39
	Yield (%)	61.47	76.31	83.12	86.42	89.12
Uttar Dinajpur	Independent variables (%)	27.12	40.24	61.55	72.43	78.61
	Yield (%)	72.52	87.14	89.23	90.86	92.26

plot indicated that selecting the second component increased the error, making it a non-conservative choice. Therefore, only the first component was selected.

2. *Jalpaiguri and Uttar Dinajpur*: Lowest *RMSE-CV* (2.82 and 3.13) at the fourth component, as explained, was 58.46 and 72.43% of independent variable variability and 94.30 and 90.86% of yield variability.
3. *Malda*: Lowest *RMSE-CV* = 6.10 at fifth component, as explained, was 86.39% of independent variable variability and 89.12% of yield variability.

The contribution of each index to the optimal PLSR components is illustrated through coefficient plots in Figure 3. The strong positive loading of the time trend

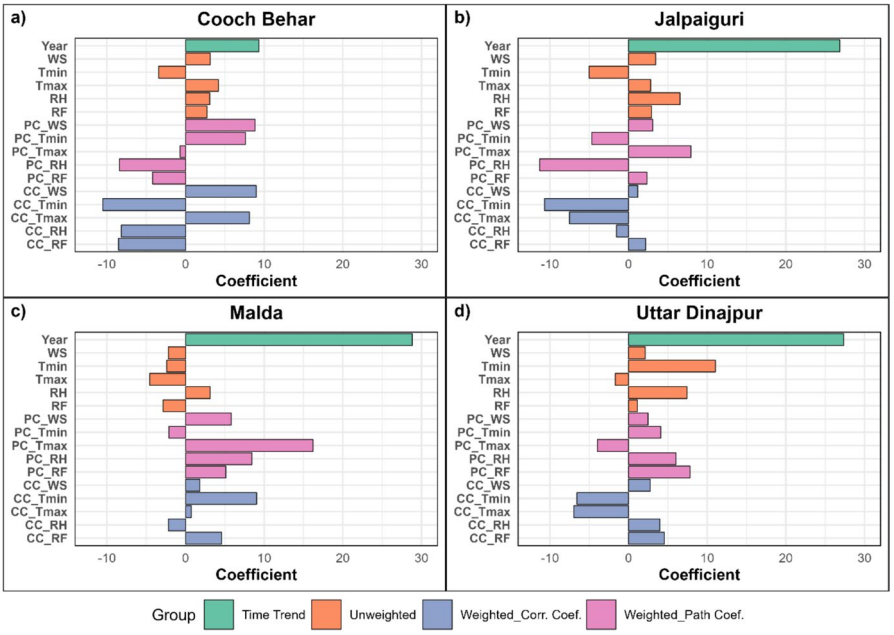

Fig. 3 Coefficient plots of PLSR indicating the importance of indices

variable (Year) across all districts was due to a significant positive trend in potato yield. In Jalpaiguri, Malda and Uttar Dinajpur, path coefficient-based indices contributed more than correlation-based indices, while in Cooch Behar, both indices contributed equally.

Model Fitting

Artificial Neural Network Model

Three ANN models were developed for each district using SR-selected variables, principal component scores and PLSR components as inputs. The optimal number of hidden-layer neurons was determined via tenfold cross-validation, selecting the configuration with the lowest *RMSE-CV* (Supplementary file 3). The results of the ANN model with three activation functions and optimized parameters for all four districts are summarized in Fig. 4.

1. Cooch Behar: Four hidden neurons for SR_ANN and PCA_ANN and one for PLSR_ANN were found to be optimal. Tanh achieved the lowest *SSE* (0.03, 0.03 and 0.25) with learning rates 0.01, 0.01 and 0.07 and showed the best performance with low *AIC*.

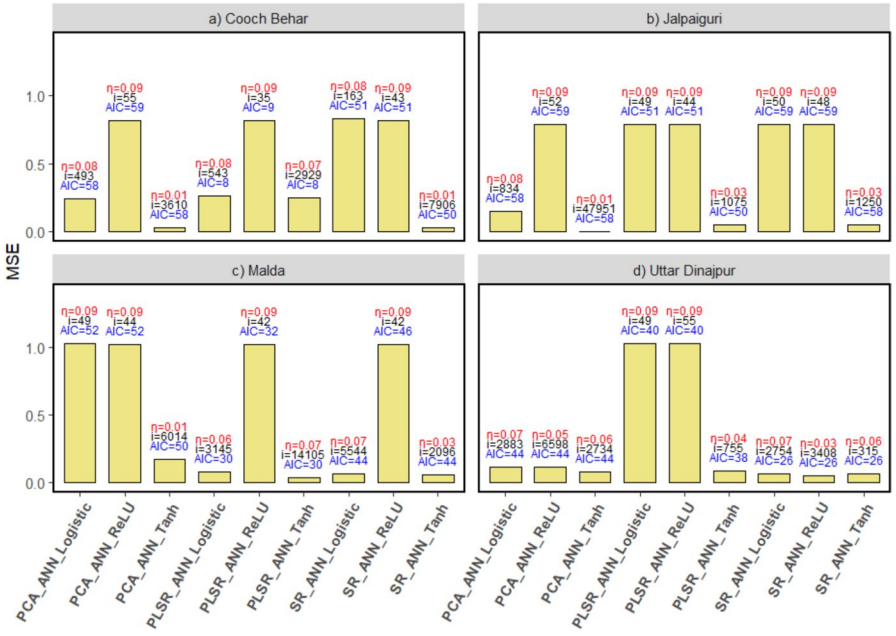

Fig. 4 Results of ANN models with three activation functions indicating optimized learning rate (η), number of iterations (i) and *AIC*

2. Jalpaiguri: Four hidden neurons were optimal for SR_ANN and PCA_ANN, and it is two for PLSR_ANN. Tanh converged to the lowest error with a 0.01 and 0.03 learning rate, taking more steps but ensured the lowest *AIC*.
3. Malda: Optimal hidden neurons: 3 (SR-ANN), 4 (PCA-ANN) and 2 (PLSR_ANN). Tanh performed best with lowest *SSE* (0.06, 0.17 and 0.04) and learning rates (0.03, 0.01 and 0.07).
4. Uttar Dinajpur: 2 hidden neurons (SR_ANN), 3 (PCA_ANN and PLSR_ANN). ReLU (SR_ANN) had the lowest *SSE* (0.03, learning rate 0.05), while Tanh (PCA_ANN, PLSR_ANN) performed best with *SSE* (0.08, 0.09) and low *AIC* with learning rates 0.06 and 0.04.

The Tanh activation function outperformed other functions in all models except for SR_ANN in Uttar Dinajpur, where ReLU was optimal. Tanh is particularly effective for yield forecasting due to its zero-centred output, unlike the sigmoid function, which accelerates learning (Nwankpa et al. 2018). Its ability to serve as a robust nonlinear function in the hidden layer further enhances model performance (Feng and Lu 2019).

The optimal learning rate ranged from 0.01 to 0.09, ensuring the convergence to the lowest possible error in the backpropagation algorithm. A low learning rate slowed convergence, while a high rate risks settling at a local optimum instead of the global one. The selection of an optimal rate prioritizes error minimization, even at the relatively lowest learning rate (Kandil et al.1993). These findings align with the results of those of Igiri et al. (2015).

The number of hidden-layer neurons ranged from one to five, consistent with the findings of Karmakar and Goswami (2021). Additionally, the number of hidden neurons was lower than the input factors, aligning with the results of Silaban and Zarlis (2017).

The SR_ANN, PCA_ANN and PLSR_ANN models, each with its optimal activation function, were selected for the final comparison with other models.

Support Vector Regression Model

Three SVR models were developed for each district to predict potato yield using SR-selected variables, PC scores and PLSR components as inputs. The performance of linear, polynomial and radial basis function (*RBF*) kernels was evaluated. C was set at a range from 1 to 32, ϵ from 0 to 1 and γ from 0 to 1. Degree parameter (d) in polynomial kernels tuned from 1 to 10. The hyperparameters are optimized using a grid search with tenfold cross-validation and are depicted in Fig. 5.

1. *Cooch Behar*: RBF kernel performed best with ($C=32$, $\epsilon=0$ and $\gamma=0.1$), ($C=4$, $\epsilon=0$ and $\gamma=0.3$) and ($C=8$, $\epsilon=0.3$ and $\gamma=0.1$), yielded the lowest *RMSE* in SR_SVR, PCA_SVR and PLSR_SVR models, respectively. A γ value greater than zero indicates slight to moderate nonlinearity in the model that can capture a non-linear relationship between the input and yield.

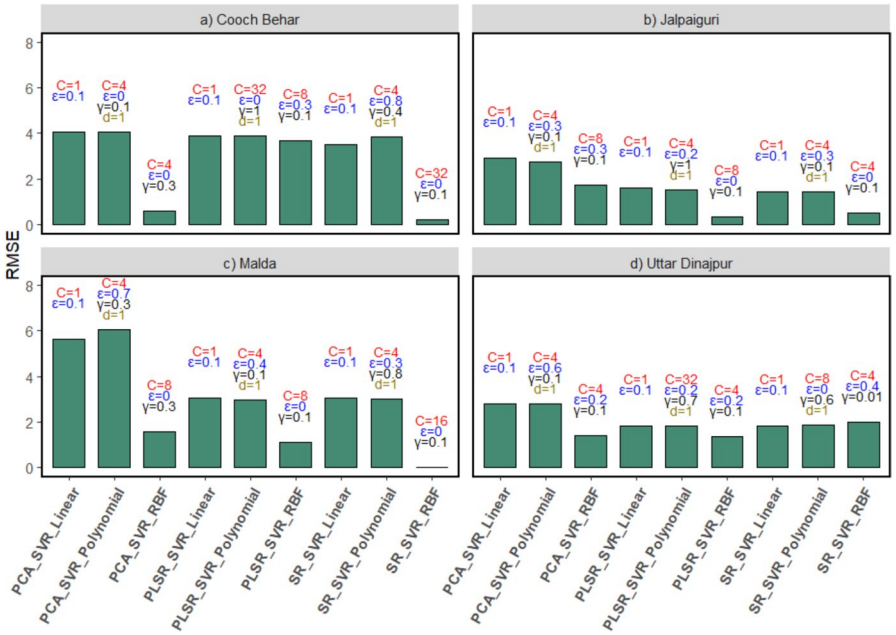

Fig. 5 Results of SVR models with three kernel functions indicating optimized hyperparameters: cost (*C*), epsilon (ϵ), gamma (*Y*) and degree parameter (*d*)

2. *Jalpaiguri*: *RBF* kernel performed best with ($C=4$, $\epsilon=0$ and $\gamma=0.1$), ($C=8$, $\epsilon=0.3$ and $\gamma=0.1$) and ($C=8$, $\epsilon=0$ and $\gamma=0.1$) yielding the lowest *RMSE* in SR_SVR, PCA_SVR and PLSR_SVR models, respectively. A γ value of 0.1 indicates slight nonlinearity in the model, resulting from the nonlinear relationship between the input and yield.
3. *Malda*: The lowest *RMSE* was achieved by *RBF* kernels in all three models. There existed a moderate nonlinearity in the SR_SVR model and a slight nonlinearity in PCA_SVR and PLSR_SVR model.
4. *Uttar Dinajpur*: Linear kernel in SR_SVR had the lowest *RMSE*, while nonlinear *RBF* performed best in PCA_SVR and PLSR_SVR with slight nonlinearity ($\gamma=0.1$).

The *RBF* kernel consistently produced the lowest *RMSE* across models, making it more effective for yield prediction (Shafiee et al. 2021; Truong and Pham 2021). The linear kernel had a narrow hyperplane margin (ϵ), and a very low ϵ increased the number of support vectors, leading to overfitting. In contrast, the *RBF* kernel maintained a balanced margin width. A higher cost resulted in a tighter hyperplane margin, reducing the number of support vectors (Guenther and Schonlau 2016). Hyperparameter-tuned SVR models demonstrated efficient performance (Saha et al. 2021; Sun et al. 2021; Ozden and Guleryuz 2022).

The SR_SVR, PCA_SVR and PLSR_AVR models, each with their optimal kernel function and hyperparameters, were selected for the final comparison with other models.

Penalized Regression Models

The optimal penalty parameters for penalized regression models for potato yield across districts are optimized using the cross-validation procedure, and the same are summarized in Table 5. The *LASSO* lambda (λ) value was one for all districts, while *RR* had higher λ values except in Malda, where the *LASSO* penalty exceeds *RR*. *ENET* models in Cooch Behar and Uttar Dinajpur, with λ values below 0.5, resembled Ridge regression, whereas in Malda, the *ENET* model was identical to *LASSO* ($\alpha = 1$; λ is equal to *LASSO*).

Penalized regression models were fitted using these optimal parameters. In all districts, the sum of absolute coefficient values in penalized models was lower than in *OLS* (Supplementary file 1). In Cooch Behar, *LASSO* retained five variables, while *ENET* removed only five. The sum of absolute coefficients followed the order: *LASSO* < *ENET* < *RR*. In Jalpaiguri, RR had the lowest sum, followed by *ENET* and *LASSO*, with four and five variables retained in *LASSO* and *ENET* models, respectively. Malda's *LASSO* and *ENET* models yielded identical coefficients due to matching penalty parameters, both had lower absolute coefficient sums than *RR*.

Evaluation of Goodness of Fit and Validation of Fitted Models

The performance of fitted models was assessed using R^2, *MAE*, *RMSE* and *nRMSE*. *RMSE* values for training and validation are shown in Fig. 6, while other evaluation metrics are in Table 6.

1. *Cooch Behar*: SR_SVR, PCA-ANN and SR_ANN models had high R^2 with low *MAE*, *RMSE* and *nRMSE* for training, but significantly lower testing errors indicated potential overfitting. The PLSR_ANN model ($R^2 = 0.79$) exhibited the lowest training errors (*MAE* 1.98, *RMSE* 1.64 and *nRMSE* 0.13) and testing errors (*MAE* 1.56, *RMSE* 1.79 and *nRMSE* 0.10). Figure 6 shows consistently low *RMSE* for this model during both calibration and validation. Overall, PLSR_ANN was the most effective model for potato yield prediction in Cooch Behar.

Table 5 Optimum penalty parameters of penalized regression models

District	Ridge	*LASSO*	Elastic Net	
	λ	λ	α	λ
Cooch Behar	6	1	0.1	4
Jalpaiguri	4	1	0.6	1
Malda	3	1	1.0	1
Uttar Dinajpur	2	1	0.4	1

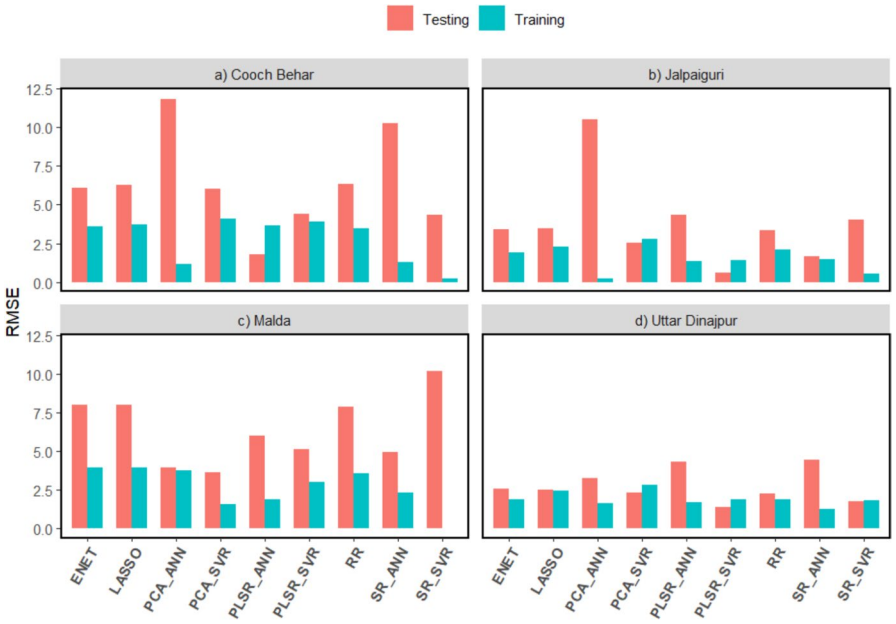

Fig. 6 *RMSE* of all the fitted models in both training and testing stage

2. *Jalpaiguri*: The PCA-ANN model had the highest R^2 (0.99) but exhibited high testing errors. PLSR_SVR ($R^2 = 0.92$) achieved lower errors in both training (*MAE* 1.14, *RMSE* 1.44, *nRMSE* 0.06) and testing (*MAE* 0.52, *RMSE* 0.61 and *nRMSE* 0.03). Figure 6 indicates PLSR_SVR model has consistently low *RMSE* during training and validation. SR_ANN model also performed well. The PLSR_SVR model was the most reliable for potato yield prediction in Jalpaiguri followed by SR_ANN.
3. *Malda*: SR_SVR model achieved the highest R^2 (0.99) with zero training errors but exhibited high testing errors. The PCA_SVR model ($R^2 = 0.97$) had low training errors (*MAE* 0.35, *RMSE* 1.56 and *nRMSE* 0.06) but higher testing errors (*MAE* 3.36, *RMSE* 3.65 and *nRMSE* 0.15). Figure 6 shows that PCA_SVR maintains low *RMSE* during calibration and validation. Other models, including SR_ANN, PCA-ANN and PLSR_ANN also had high R^2 with low errors in both stages. Overall, PCA_SVR was the best-performing model for potato yield prediction in Malda, followed by SR_ANN, PCA_ANN and PLSR_ANN
4. *Uttar Dinajpur*: PLSR_SVR and SR_SVR models had R^2 around 0.90. Their training errors were similar, but the testing errors of SR_SVR were comparatively higher. The PLSR_SVR model ($R^2 = 0.90$) demonstrated the lowest testing errors (*MAE* 1.13, *RMSE* 1.39 and *nRMSE* 0.06). Figure 6 confirms the low *RMSE* during both calibration and validation of PLSR_SVR. Thus, PLSR_SVR was the best-performing model for Uttar Dinajpur, followed by SR_SVR.

Table 6 The model evaluation parameters for training and testing stages of fitted models for potato crop

Model	Training			Testing		Training			Testing	
	R^2	*MAE*	*nRMSE*	*MAE*	*nRMSE*	R^2	*MAE*	*nRMSE*	*MAE*	*nRMSE*
	Cooch Behar					Jalpaiguri				
SR_ANN	0.96	1.14	0.06	8.95	0.55	0.93	1.20	0.07	0.53	0.04
SR_SVR	0.99	0.05	0.01	4.14	0.21	0.99	0.20	0.02	3.63	0.19
PCA_ANN	0.97	0.94	0.05	11.28	0.73	0.99	0.13	0.01	14.21	0.99
PCA_SVR	0.62	2.80	0.17	4.44	0.30	0.75	2.15	0.11	2.55	0.12
PLSR_ANN	0.79	1.98	0.13	1.56	0.10	0.94	1.09	0.07	4.28	0.25
PLSR_SVR	0.64	2.83	0.16	3.36	0.22	0.92	1.14	0.06	0.52	0.03
RR	0.72	2.69	0.14	0.59	0.31	0.86	1.55	0.08	0.88	0.16
LASSO	0.67	2.96	0.15	0.52	0.31	0.83	1.87	0.09	0.87	0.17
ENET	0.70	2.79	0.15	0.57	0.30	0.88	1.51	0.08	0.90	0.16
	Malda					Uttar Dinajpur				
SR_ANN	0.94	1.83	0.08	4.56	0.21	0.95	0.98	0.07	4.13	0.32
SR_SVR	0.99	0.00	0.00	7.25	0.43	0.90	1.43	0.08	1.34	0.08
PCA_ANN	0.84	2.66	0.13	3.02	0.17	0.92	1.20	0.09	2.69	0.23
PCA_SVR	0.97	0.35	0.06	3.36	0.15	0.76	2.62	0.12	1.94	0.10
PLSR_ANN	0.96	1.44	0.07	5.52	0.25	0.91	1.36	0.10	3.00	0.31
PLSR_SVR	0.90	2.51	0.11	4.23	0.22	0.90	1.52	0.08	1.13	0.06
RR	0.85	2.82	0.13	0.59	0.33	0.89	1.61	0.08	0.90	0.10
LASSO	0.82	2.93	0.15	0.59	0.34	0.82	2.02	0.10	0.82	0.11
ENET	0.82	2.93	0.15	0.59	0.34	0.90	1.54	0.08	0.83	0.12

Overall, PLSR_SVR and PLSR_ANN were most effective models for potato yield prediction across districts. Since PLSR components were derived by considering both yield and weather indices, they enhanced model predictability by retaining the most relevant factors influencing yield in the derived PLSR components. Figure 3 highlights the significant contribution of path coefficient-based indices in PLSR component development, improving the model's ability to capture potato yield variability effectively. Critical weather parameters, particularly maximum temperature and relative humidity, significantly affect potato yield by increasing the risk of late blight disease (Joshi et al. 2025). Temperature and rainfall across the crop growing season greatly influence the potato yield (Khan et al. 2024). Temperature anomalies significantly influence potato yield, with optimal temperature requirements varying across different crop growth stages (Biswal et al. 2025; Naz et al. 2025). Integrating the weighted influence of weekly weather parameters on yield through weighted weather indices significantly improved the model's predictive accuracy. Accurately quantifying regional climate vulnerability and its impact is a critical step toward developing reliable early potato yield forecasting systems in India (Shah et al. 2023). Assigning appropriate weights in the composite index reduced bias in regional-level climate-based models (Dhamija et al. 2020). PCA_ANN performed well during calibration but poorly during validation, indicating potential overfitting, consistent with findings by Das et al. (2018) and Aravind et al. (2022). This challenges the

assumption that principal components explaining the most variance in independent variables have high predictive power for the response variable (Das et al. 2020). While components explaining less variance are often excluded, they can sometimes be crucial for prediction (Kung and Sharif 1980; Jolliffe 1982). PLSR, however, derived components that maximize covariance between dependent and independent variables, enhanced predictability (Hu et al. 2018; Duan et al.2020). With just one or two components, PLSR efficiently explained the response variable, requiring fewer components than PCA (Wentzell and Montoto 2003). For predictive modeling, PLSR was generally more effective than PCA (Carrascal et al. 2009; Mahesh et al. 2015; Wang and Li 2020; Liu et al. 2022). ANN with PLSR components performed well in predictive contexts (Sharabiani et al. 2023). Machine learning models can identify complex patterns that regression models often miss (Matsumura et al. 2015). Nonlinear machine learning models, such as SVR and ANN, outperformed linear regression, highlighting the nonlinear influence of weather factors on potato yield (Crane-Droesch 2018; Bouras et al. 2021; Ajith et al. 2023b; Sridhara et al. 2023).

Conclusion

Incorporating the weighted influence of weekly weather through weighted weather indices significantly enhanced the predictive accuracy of the models. The weather indices, particularly the proposed optimally weighted path coefficient-based indices, served as highly informative inputs, contributing to optimized model performance by effectively condensing weekly weather data and capturing the critical weather-potato yield relationships. Among feature selection methods, PLSR-based selection enhanced predictability by retaining key yield-influencing factors. Path coefficient-based indices which effectively captured potato yield variability contributed significantly to PLSR component development. For modeling, ANN with a tangent hyperbolic activation function, a learning rate of 0.01–0.09 and hidden layer neurons fewer than input layers performed best. SVR with a radial basis function kernel and optimized hyperparameters effectively captured slight to moderate nonlinear relationships between weather factors and potato yield. Overall, fine-tuned nonlinear machine learning models, ANN and SVR using PLSR-based inputs with proper hyperparameter optimization were efficient in location-specific potato yield predictions by capturing complex weather-yield interactions. This systematic approach, beginning from the construction of predictive and information-rich input features such as composite weather indices that optimally weigh the effect of weekly weather conditions on yield, PLSR-based dimension reduction and fine-tuning of hyperparameters of ML models, is vital for effectively capturing complex nonlinear associations between weather and yield. Such a comprehensive approach is crucial for building robust, region-specific potato yield predictive models with enhanced accuracy.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11540-025-09924-3.

Acknowledgements We extend our gratitude to the Department of Agricultural Statistics in Uttar Banga Krishi Viswavidyalaya, Cooch Behar, West Bengal, India, for the technical support.

Author Contribution Ajith S: conceptualization; investigation; methodology; formal analysis; visualization and writing — original draft; M. K. Debnath: conceptualization, supervision and validation; D. S. Gupta and P. S. Patra: supervision and validation; S. Bandyopadhyay: meteorological data supply; D. Sarkar and Rakesh S: review and editing.

Data Availability Crop yield data can be availed from the official website of Directorate of Economics and Statistics (DES), Ministry of Agriculture & Farmers Welfare, Government of India (https://data.desagri.gov.in/website/crops-apy-report-web). The authors are not authorized to distribute weather data obtained from Indian Meteorological Department.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdiscip Rev: Comput Stat 2(1):97–106. https://doi.org/10.1002/wics.51

Agrawal R, Mehta SC (2007) Weather based forecasting of crop yields, pests and diseases-IASRI models. J Ind Soc Agril Statist 61(2):255–263

Agrawal R, Jain RC, Jha MP (1986) Models for studying rice crop-weather relationship. Mausam 37(1):67–70. https://doi.org/10.54302/mausam.v37i1.2153

Ajith S, Debnath MK, Gupta DS, Basak P, Bandyopadhyay S, Kheroar S, Ragini HR (2023) Comparative evaluation of penalized regression models with multiple linear regression for predicting rapeseed-mustard yield: weather-indices based approach. J Agromet 25(3):432–439. https://doi.org/10.54386/jam.v25i3.2185

Ajith S, Debnath MK, Gupta DS, Basak P (2023b) Application of statistical and machine learning models in combination with stepwise regression for predicting rapeseed-mustard yield in Northern districts of West Bengal. Int J Stat Appl Math 8(3):141–149. https://doi.org/10.22271/maths.2023.v8.i3b.1004

Ajith S, Debnath MK, Karthik R (2024) Statistical and machine learning models for location-specific crop yield prediction using weather indices. Int J Biometeorol 68:2453–2475. https://doi.org/10.1007/s00484-024-02763-w

Ajith S, Vijayakumar S, Elakkiya N (2025) Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms. Dis Food 5:67. https://doi.org/10.1007/s44187-025-00338-1

Al Azies H, Trishnanti D, PH EM (2019) Comparison of kernel support vector machine (SVM) in classification of human development index (HDI). IPTEK J Proc Ser (6):53–57. https://doi.org/10.12962/j23546026.y2019i6.6394

Alwin DF, Hauser RM (1975) The decomposition of effects in path analysis. Am Sociol Rev 40:37–47

Ansarifar J, Wang L, Archontoulis SV (2021) An interaction regression model for crop yield prediction. Sci Rep 11(1):17754. https://doi.org/10.1038/s41598-021-97221-7

Aravind KS, Vashisth A, Krishanan P, Das B (2022) Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. J Agromet 24(1):18–25. https://doi.org/10.54386/jam.v24i1.1002

Azizabadi EC, Badreldin N (2025) A review on potato crop yield and nitrogen management utilizing remote/proximal sensing technologies and machine learning models in Canada. Potato Res 68:1659–1679. https://doi.org/10.1007/s11540-024-09803-3

Basha SM, Rajput DS, Janet J, Somula RS, Ram S (2020) Principles and practices of making agriculture sustainable: crop yield prediction using Random Forest. Scalable Comput Pract Exp 21(4):591–599. https://doi.org/10.12694/scpe.v21i4.1714

Batool D, Shahbaz M, Shahzad Asif H, Shaukat K, Alam TM, Hameed IA, Luo S (2022) A hybrid approach to tea crop yield prediction using simulation models and machine learning. Plants 11(15):1925. https://doi.org/10.3390/plants11151925

Bender FD, Sentelhas PC (2020) Assessment of regional climate change impacts on Brazilian potato tuber yield. Int J Plant Prod 14(4):647–661. https://doi.org/10.1007/s42106-020-00111-7

Bhardwaj M, Kumar P, Kumar S, Dagar V, Kumar A (2022) A district-level analysis for measuring the effects of climate change on production of agricultural crops, i.e. wheat and paddy: evidence from India. Env Sci Poll Res 29(21):31861–31885. https://doi.org/10.1007/s11356-021-17994-2

Bhat S, Vidya R (2018) Partial least squares and other biased regression methods. Thail Stat 16(1):38–55

Biswal P, Swain DK, Jha MK (2025) Understanding the significance of weather and soil microclimate for improvement of potato yield using SUBSTOR and statistical models. Potato Res 68:1659–1679. https://doi.org/10.1007/s11540-024-09803-3

Bouras EH, Jarlan L, Er-Raki S, Balaghi R, Amazirh A, Richard B, Khabba S (2021) Cereal yield forecasting with satellite drought-based indices, weather data and regional climate indices using machine learning in Morocco. Remote Sens 13(16):3101. https://doi.org/10.3390/rs13163101

Carrascal LM, Galván I, Gordo O (2009) Partial least squares regression as an alternative to current regression methods used in ecology. Oikos 118(5):681–690. https://doi.org/10.1111/j.1600-0706.2008.16881.x

Crane-Droesch A (2018) Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ Res Lett 13(11):114003. https://doi.org/10.1088/1748-9326/aae159

Daoud JI (2017) Multicollinearity and regression analysis. In J Phys Conf Ser 949(1):012009. https://doi.org/10.1088/1742-6596/949/1/012009

Das B, Nair B, Reddy VK, Venkatesh P (2018) Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. Int J Biometeorol 62(10):1809–1822. https://doi.org/10.1007/s00484-018-1583-6

Das B, Nair B, Arunachalam V, Reddy KV, Venkatesh P, Chakraborty D, Desai S (2020) Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the west coast of India. Int J Biometeorol 64:1111–1123. https://doi.org/10.1007/s00484-020-01884-2

Dhamija V, Shukla R, Gornott C, Joshi PK (2020) Consistency in vulnerability assessments of wheat to climate change-a district-level analysis in India. Sustainability 12(19):8256. https://doi.org/10.3390/su12198256

Duan L, Xie H, Li Z, Yuan H, Guo Y, Xiao X, Zhou Q (2020) Use of partial least squares regression to identify factors controlling rice yield in Southern China. Agron J 112(3):1502–1516. https://doi.org/10.1002/agj2.20161

Feng J, Lu S (2019) Performance analysis of various activation functions in artificial neural networks. In J Phys Conf Ser 1237(2):022030. https://doi.org/10.1088/1742-6596/1237/2/022030

Firinguetti L, Kibria G, Araya R (2017) Study of partial least squares and ridge regression methods. Commun Stat Simul Comput 46(8):6631–6644. https://doi.org/10.1080/03610918.2016.1210168

Freedman DA (1987) As others see us: a case study in path analysis. J Educ Stat 12(2):101–128. https://doi.org/10.3102/10769986012002101

Guenther N, Schonlau M (2016) Support vector machines. Stata J 16(4):917–937. https://doi.org/10.1177/1536867X1601600407

Gupta S, Singh A, Kumar A, Shahi UP, Sinha NK, Roy S (2018) Yield forecasting of wheat and mustard for western Uttar Pradesh using statistical model. J Agromet 20(1):66–68. https://doi.org/10.54386/jam.v20i1.508

Hamed KH, Rao AR (1998) A modified Mann-Kendall trend test for autocorrelated data. J Hydrol 204(1–4):182–196. https://doi.org/10.1016/S0022-1694(97)00125-X

Hao Z, Geng X, Wang F, Zheng J (2018) Impacts of climate change on agrometeorological indices at winter wheat overwintering stage in Northern China during 2021–2050. Int J Climatol 38(15):5576–5588. https://doi.org/10.1002/joc.5764

Hasan MAM, Xu S, Kabir MMJ, Ahmad S (2016) Performance evaluation of different kernels for support vector machine used in intrusion detection system. Int J Comp Net Com 8(6):39–53. https://doi.org/10.5121/ijcnc.2016.8604

Haverkort AJ (1987) Climate and potato crop growth interactions in Africa's continental divide region. Agromet Potato Crop 9:137–148. https://doi.org/10.17660/ActaHortic.1988.214.9

Haverkort AJ, Verhagen A (2008) Climate change and its repercussions for the potato supply chain. Potato Res 51:223–237. https://doi.org/10.1007/s11540-008-9107-0

Henderson DA, Denison DR (1989) Stepwise regression in social and psychological research. Psychol Rep 64(1):251–257. https://doi.org/10.2466/pr0.1989.64.1.251
Hope K (1971) Path analysis: supplementary procedures. Sociology 5(2):225–241. https://doi.org/10.1177/003803857100500205
Hoskuldsson A (1998) PLS regression methods. J Chemometrics 2(3):211–228. https://doi.org/10.1002/cem.1180020306
Hotelling H (1957) The relations of the newer multivariate statistical methods to factor analysis. Brit J Stat Psychol 10(2):69–79. https://doi.org/10.1111/j.2044-8317.1957.tb00179.x
Hu Y, Wei X, Hao M, Fu W, Zhao J, Wang Z (2018) Partial least squares regression for determining factors controlling winter wheat yield. Agron J 110(1):281–292. https://doi.org/10.2134/agronj2017.02.0108
Hunter D, Yu H, Pukish MS, Kolbusz J, Wilamowski BM (2012) Selection of proper neural network sizes and architectures—a comparative study. IEEE Trans Ind Inform 8(2):228–240. https://doi.org/10.1109/TII.2012.2187914
Hussain N, Sarfraz S, Javed S (2021) A systematic review on crop-yield prediction through unmanned aerial vehicles. 16th Int Conf Emerg Technol (ICET). IEEE, pp 1–9. https://doi.org/10.1109/ICET54505.2021.9689838
Huzsvai L, Zsembeli J, Kovács E, Juhász C (2022) Response of winter wheat (Triticum aestivum L.) yield to the increasing weather fluctuations in a continental region of four-season climate. Agronomy 12(2):314. https://doi.org/10.3390/agronomy12020314
Igiri CP, Anyama OU, Silas AI (2015) Effect of learning rate on artificial neural network in machine learning. Int J Eng Res Tech 4(2):359–362
Jain RC, Agrawal R, Jha MP (1980) Effect of climatic variables on rice yield and its forecast. Mausam 31(4):591–596. https://doi.org/10.54302/mausam.v31i4.3477
Jégo G, Crépeau M, Jing Q, Grant B, Smith W, Mesbah M, Qian B (2025) Potato yield projections under climate change in Canada. Agron Jour 117(1):e70017. https://doi.org/10.1002/agj2.70017
Jia W, Zhao D, Tang Y, Hu C, Zhao Y (2014) An optimized classification algorithm by neural network ensemble based on PLS and OLS. Math Probl Eng 2014(1):395263. https://doi.org/10.1155/2014/395263
Jia Q, Li M, Dou X (2022) Climate change affects crop production potential in semi-arid regions: a case study in Dingxi, Northwest China, in recent 30 years. Sustainability 14(6):3578. https://doi.org/10.3390/su14063578
Jolliffe IT (1982) A note on the use of principal components in regression. J Roy Stat Soc Ser C: Appl Stat 31(3):300–303. https://doi.org/10.2307/2348005
Jolliffe IT, Trendafilov NT, Uddin M (2003) A modified principal component technique based on the LASSO. J Comp Grap Stat 12(3):531–547. https://doi.org/10.1198/1061860032148
Joshi D, Krishnan P, Vashisth A, Kundu M, Rani A, Bag TK (2025) AI-based machine learning and multiple linear regression approach to simulate the effect of weather on the crop age at first appearance of potato late blight (*Phytophthora infestans* (mont.) de bary) disease. Potato Res 68:1825–1848. https://doi.org/10.1007/s11540-024-09795-0
Kaiser HF (1960) The application of electronic computers to factor analysis. Edu Psychol Meas 20(1):141–151. https://doi.org/10.1177/001316446002000116
Kandil N, Khorasani K, Patel RV, Sood VK (1993) Optimum learning rate for backpropagation neural networks. Proc Can Conf Electr Comput Eng. IEEE, pp 465–468. https://doi.org/10.1109/CCECE.1993.332193
Kang Y, Ozdogan M, Zhu X, Ye Z, Hain C, Anderson M (2020) Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. Environ Res Let 15(6):064005. https://doi.org/10.1088/1748-9326/ab7df9
Karmakar S, Goswami S (2021) A new procedure for optimization of hidden layer neurons during learning through gradient descent process of neural network and improvement of performance in the chaos forecasting. Iran J Comp Sci 4:293–303. https://doi.org/10.1007/s42044-021-00089-z
Khan MS, Hoogenboom G, Gillani SM, Shah AS, Khan I (2024) Effects of planting date and genotype on potato growth and yield determination in a sub-tropical continental growing environment. Potato Res. https://doi.org/10.1007/s11540-024-09833-x
Kittichotsatsawat Y, Tippayawong N, Tippayawong KY (2022) Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. Sci Rep 12(1):14488. https://doi.org/10.1038/s41598-022-18635-5

Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011) Partial least squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage 56(2):455–475. https://doi.org/10.1016/j.neuroimage.2010.07.034

Król A (2017) The application of partial least squares method in hedonic modelling. Arch Data Sci Ser A 2:1–13. https://doi.org/10.5445/KSP/1000058749/0

Kross A, Znoj E, Callegari D, Kaur G, Sunohara M, Lapen DR, McNairn H (2020) Using artificial neural networks and remotely sensed data to evaluate the relative importance of variables for prediction of within-field corn and soybean yields. Remote Sens 12(14):2230. https://doi.org/10.3390/rs12142230

Kumar S, Attri SD, Singh KK (2019) Comparison of Lasso and stepwise regression technique for wheat yield prediction. J Agromet 21(2):188–192. https://doi.org/10.54386/jam.v21i2.231

Kung EC, Sharif TA (1980) Regression forecasting of the onset of the Indian summer monsoon with antecedent upper air conditions. J Appl Met Climatol 19(4):370–380. https://doi.org/10.1175/1520-0450(1980)019%3c0370:RFOTOO%3e2.0.CO;2

Kuriqi A, Ali R, Pham QB, Montenegro Gambini J, Gupta V, Malik A, Linh NT, Joshi Y, Anh DT, Nam VT, Dong X (2020) Seasonality shift and streamflow flow variability trends in central India. Acta Geophys 68:1461–1475. https://doi.org/10.1007/s11600-020-00475-4

Lewis-Beck MS (1978) Stepwise regression: a caution. Polit Method 1:213–240. https://www.jstor.org/stable/25791533. Accessed 31 Jan 2025

Li CC (1956) The concept of path coefficient and its impact on population genetics. Biometrics 12(2):190–210. https://doi.org/10.2307/3001760

Li Q, Zhang S (2020) Impacts of recent climate change on potato yields at a provincial scale in Northwest China. Agronomy 10(3):426. https://doi.org/10.3390/agronomy10030426

Li Z, Ding L, Xu D (2022) Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. Sci Total Environ 815:152880. https://doi.org/10.1016/j.scitotenv.2021.152880

Lin Y, Li S, Duan S, Ye Y, Li B, Li G, Lyv D, Jin L, Bian C, Liu J (2023) Methodological evolution of potato yield prediction: a comprehensive review. Front Plant Sci 14:1214006. https://doi.org/10.3389/fpls.2023.1214006

Liu C, Zhang X, Nguyen TT, Liu J, Wu T, Lee E, Tu XM (2022) Partial least squares regression and principal component analysis: similarity and differences between two popular variable reduction approaches. Gen Psychiatry 35(1):e100662. https://doi.org/10.1136/gpsych-2021-100662

Lobell D, Cahill K, Field C (2006) Weather-based yield forecasts developed for 12 California crops. Calif Agric 60(4):211–215. https://doi.org/10.3733/ca.v060n04p211

Mahesh S, Jayas DS, Paliwal J, White NDG (2015) Comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using the near-infrared (NIR) hyperspectral images of bulk samples of Canadian wheat. Food Bioprocess Technol 8:31–40. https://doi.org/10.1007/s11947-014-1381-z

Matsumura K, Gaitan CF, Sugimoto K, Cannon AJ, Hsieh WW (2015) Maize yield forecasting by linear regression and artificial neural networks in Jilin. China J Agric Sci 153(3):399–410. https://doi.org/10.1017/S0021859614000392

Mavromatis T (2016) Spatial resolution effects on crop yield forecasts: an application to rainfed wheat yield in north Greece with CERES-Wheat. Agric Syst 143:38–48. https://doi.org/10.1016/j.agsy.2015.12.002

Mukiibi A, Machakaire AT, Franke AC, Steyn JM (2025) A systematic review of vegetation indices for potato growth monitoring and tuber yield prediction from remote sensing. Potato Res 68(1):409–448. https://doi.org/10.1007/s11540-024-09748-7

Naz S, Ahmed M, Abbas G, Fatima Z, Hussain S, Iqbal P, Ghani A, Ali M, Awan TH, Samad N, Aasim M (2025) Assessment of climate change impact on potato-potato cropping system under semi-arid environment and designing of adaptation strategies. Potato Res 68:1209–1239. https://doi.org/10.1007/s11540-024-09771-8

Niles HE (1923) The method of path coefficients an answer to wright. Genetics 8(3):256. https://doi.org/10.1093/genetics/8.3.256

Nti IK, Nyarko-Boateng O, Adekoya FA, Weyori BA (2021) An empirical assessment of different kernel functions on the performance of support vector machines. Bul Elect Eng Inform 10(6):3403–3411. https://doi.org/10.11591/eei.v10i6.3046

Nwankpa C, Ijomah W, Gachagan A, Marshall S (2018) Activation functions: comparison of trends in practice and research for deep learning. arXiv preprint. https://doi.org/10.48550/arXiv.1811.03378

Olusegun AM, Dikko HG, Gulumbe SU (2015) Identifying the limitation of stepwise selection for variable selection in regression analysis. Am J Theor Appl Stat 4(5):414–419

Ozden E, Guleryuz D (2022) Optimized machine learning algorithms for investigating the relationship between economic development and human capital. Comput Econ 60(1):347–373. https://doi.org/10.1007/s10614-021-10194-7

Paidipati KK, Chesneau C, Nayana BM, Kumar KR, Polisetty K, Kurangi C (2021) Prediction of rice cultivation in India—support vector regression approach with various kernels for non-linear patterns. Agri Engineering 3(2):182–198. https://doi.org/10.3390/agriengineering3020012

Paudel D, Boogaard H, de Wit A, Janssen S, Osinga S, Pylianidis C, Athanasiadis IN (2021) Machine learning for large-scale crop yield forecasting. Agric Syst 187:103016. https://doi.org/10.1016/j.agsy.2020.103016

Pope PT, Webster JT (1972) The use of an F-statistic in stepwise regression procedures. Technometrics 14(2):327–340. https://doi.org/10.1080/00401706.1972.10488919

Puig-Arnavat M, Bruno JC (2015) Artificial neural networks for thermochemical conversion of biomass. In: Recent advances in thermo-chemical conversion of biomass. Elsevier, Amsterdam, pp 133–156. https://doi.org/10.1016/B978-0-444-63289-0.00005-3

Qu H, Zhang Y (2016) A new kernel of support vector regression for forecasting high-frequency stock returns. Math Prob Eng 2016:4907654. https://doi.org/10.1155/2016/4907654

Rai S, Nandre J, Kanawade BR (2022) A comparative analysis of crop yield prediction using regression. In 2nd Int Conf Intell Technol (CONIT) pp 1–4. IEEE. https://doi.org/10.1109/CONIT55038.2022.9847783

Roberts MJ, Braun NO, Sinclair TR, Lobell DB, Schlenker W (2017) Comparing and combining process-based crop models and statistical models with some implications for climate change. Environ Res Let 12(9):095010. https://doi.org/10.1088/1748-9326/aa7f33

Rötter RP, Carter TR, Olesen JE, Porter JR (2011) Crop–climate models need an overhaul. Nat Clim Change 1(4):175–177. https://doi.org/10.1038/nclimate1152

Ruengvirayudh P, Brooks GP (2016) Comparing stepwise regression models to the best-subsets models, or, the art of stepwise. Gen Lin Model J 42(1):1–14

Saha A, Singh KN, Ray M, Rathod S, Choudhury S (2021) Modelling and forecasting cotton production using tuned-support vector regression. Cur Sci 121(8):1090–1098. https://doi.org/10.18520/cs/v121/i8/1090-1098

Savas C, Dovis F (2019) The impact of different kernel functions on the performance of scintillation detection based on support vector machines. Sensors 19(23):5219. https://doi.org/10.3390/s19235219

Shafiee S, Lied LM, Burud I, Dieseth JA, Alsheikh M, Lillemo M (2021) Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. Comp Elec Agric 183:106036. https://doi.org/10.1016/j.compag.2021.106036

Shah MH, Ajaharuddin SM, Kundu S, Kumawat SM, Atta K, Hossain A, Hembram A, Pramanik K, Ali MS, Islam S, Pande CB (2023) Potential impact of changing climate on the sustainability of potato (*Solanum tuberosum* l.) production in India. In climate change impacts in India. Earth and Environmental Sciences Library, Springer, Cham. https://doi.org/10.1007/978-3-031-42056-6_14

Sharabiani V, Soltani Nazarloo A, Taghinezhad E, Veza I, Szumny A, Figiel A (2023) Prediction of winter wheat leaf chlorophyll content based on VIS/NIR spectroscopy using ANN and PLSR. Food Sci Nut 11(5):2166–2175. https://doi.org/10.1002/fsn3.3071

Siebert S, Webber H, Rezaei EE (2017) Weather impacts on crop yields-searching for simple answers to a complex problem. Environ Res Lett 12(8):081001. https://doi.org/10.1088/1748-9326/aa7f15

Silaban H, Zarlis M (2017) Analysis of accuracy and epoch on back-propagation BFGS quasi-newton. In J Physics: Conf Ser 930(1):012006. https://doi.org/10.1088/1742-6596/930/1/012006

Singh BP, Dua VK, Govindakrishnan PM, Sharma S (2013) Impact of climate change on potato. In Climate-resilient horticulture: adaptation and mitigation strategies. Springer, India. https://doi.org/10.1007/978-81-322-0974-4_12

Singh A, Nagar V, Tanwar N, Shahi UP, Bhan SC (2021) Use of statistical models in yield forecasting of wheat, mustard and potato crop in western districts of Uttar Pradesh. India Int J Environ Clim Change 11(12):573–579. https://doi.org/10.9734/IJECC/2021/v11i1230648

Sridhara S, Ramesh N, Gopakkali P, Das B, Venkatappa SD, Sanjivaiah SH, Elansary HO (2020) Weather-based neural network, stepwise linear and sparse regression approach for rabi sorghum yield forecasting of Karnataka, India. Agronomy 10(11):1645. https://doi.org/10.3390/agronomy10111645

Sridhara S, Manoj KN, Gopakkali P, Kashyap GR, Das B, Singh KK, Srivastava AK (2023) Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. Int J Biometeorol 67(1):165–180. https://doi.org/10.1007/s00484-022-02396-x

Struik PC (2007) Responses of the potato plant to temperature. In Potato biology and biotechnology. Elsevier Science BV, pp 367–393. https://doi.org/10.1016/B978-044451018-1/50060-9

Sun Y, Ding S, Zhang Z, Jia W (2021) An improved grid search algorithm to optimize SVR for prediction. Soft Comput 25:5633–2544. https://doi.org/10.1007/s00500-020-05560-w

Tamayo-Vera D, Mesbah M, Zhang Y, Wang X (2025) Advanced machine learning for regional potato yield prediction: analysis of essential drivers. npj Sustain Agric 3(1):12. https://doi.org/10.1038/s44264-025-00052-6

Trnka M, Olesen JE, Kersebaum KC, Rötter RP, Brázdil R, Eitzinger J, Rajdl K (2016) Changing regional weather crop yield relationships across Europe between 1901 and 2012. Climate Res 70(2–3):195–214. https://doi.org/10.3354/cr01426

Truong VH, Pham HA (2021) Support vector machine for regression of ultimate strength of trusses: a comparative study. Engineering J 25(7):157–166. https://doi.org/10.4186/ej.2021.25.7.157

Van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: a systematic literature review. Comp Elect Agric 177:105709. https://doi.org/10.1016/j.compag.2020.105709

Vander-Velde M, Biavetti I, El-Aydam M, Niemeyer S, Santini F, van den Berg M (2019) Use and relevance of European Union crop monitoring and yield forecasts. Agric Syst 168:224–230. https://doi.org/10.1016/j.agsy.2018.05.001

Vapnik V (2000) SVM method of estimating density, conditional probability, and conditional density. In 2000 IEEE Int Symp Circ Syst (ISCAS) (2):749–752. IEEE. https://doi.org/10.1109/ISCAS.2000.856437

Vatcheva KP, Lee M, McCormick JB, Rahbar MH (2016) Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology (Sunnyvale, Calif.) 6(2). https://doi.org/10.4172/2161-1165.1000227

Vora K, Yagnik S (2014) A survey on backpropagation algorithms for feedforward neural networks. Int J Engin Develop Res 1(3):193–197

Wang W, Li H (2020) Partial least squares regression for predicting the speed of electromagnetic fuze plate. In J Phy: Conf Ser 1549(3):032084. IOP Publishing. https://doi.org/10.1088/1742-6596/1549/3/032084

Wang P, Chen S, Yang S (2022) Recent advances on penalized regression models for biological data. Mathematics 10(19):3695. https://doi.org/10.3390/math10193695

Wentzell PD, Montoto LV (2003) Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. Chemomet Intell Lab Syst 65(2):257–279. https://doi.org/10.1016/S0169-7439(02)00138-7

Wojciuk M, Swiderska-Chadaj Z, Siwek K, Gertych A (2024) Improving classification accuracy of fine-tuned CNN models: impact of hyperparameter optimization. Heliyon 10(5):e26586. https://doi.org/10.1016/j.heliyon.2024.e26586

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J Roy Stat Soc Ser b: Stat Meth 67(2):301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Authors and Affiliations

Ajith S[1] · Manoj Kanti Debnath[1] · Deb Sankar Gupta[1] · Deepranjan Sarkar[2] · Subhendu Bandyopadhyay[3] · Partha Sarathi Patra[4] · Rakesh S[5]

✉ Manoj Kanti Debnath
mkanti1984@gmail.com

Ajith S
ajithagristat@gmail.com

[1] Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Cooch Behar, West Bengal 736165, India

[2] Department of Food, Agriculture and Natural Resources, School of Environment, Resources and Development, Asian Institute of Technology, Klong Luang, Pathum Thani 12120, Thailand

[3] Department of Agronomy, Uttar Banga Krishi Viswavidyalaya, Cooch Behar, West Bengal 736165, India

[4] Regional Research Station, Terai Zone, Uttar Banga Krishi Viswavidyalaya, Cooch Behar, West Bengal 736165, India

[5] International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, Telangana 502324, India