FORECASTING YIELD OF MAJOR CROPS IN THE NORTHERN DISTRICTS OF WEST BENGAL

A thesis

submitted to the

Uttar Banga Krishi Viswavidyalaya

in partial fulfilment of the requirement

for the award of the degree of

Doctor of Philosophy (Agriculture)

in

AGRICULTURAL STATISTICS

Submitted by

AJITH S

Registration Number: A-2020-012-D

Department of Agricultural Statistics,
Faculty of Agriculture,
Uttar Banga Krishi Viswavidyalaya,
Pundibari, Cooch Behar, West Bengal-736165.

2023

UTTAR BANGA KRISHI VISWAVIDYALAYA
Pundibari, Cooch Behar, West Bengal-736165.
Department of Agricultural Statistics
FACULTY OF AGRICULTURE
www.ubkv.ac.in

From:
Dr. M.K. Debnath,
Chairman, Advisory Committee

Mobile: +91 8670937121
+91 8837481075
Email: mkanti1984@gmail.com

No.................................

Dated,.18/07/2023

Certificate

This is to certify that the work incorporated in the thesis entitled **"FORECASTING YIELD OF MAJOR CROPS IN THE NORTHERN DISTRICTS OF WEST BENGAL"** submitted by **AJITH S (Reg. No: A-2020-012-D)** in partial fulfilment of the requirements for the Doctor of Philosophy (Agriculture) in Agricultural Statistics of the Uttar Banga Krishi Viswavidyalaya, is a faithful record of bonafide research work carried out under my supervision and guidance and that no part of this thesis has been submitted for any other degree.

The assistance and help received from different sources during the course of investigation have been duly acknowledged.

18/07/2023

Place: UBKV, Pundibari
Date: 18/07/2023

(Dr. M. K. Debnath)
Chairman,
Advisory Committee

UTTAR BANGA KRISHI VISWAVIDYALAYA
Pundibari, Cooch Behar, West Bengal-736165.
Department of Agricultural Statistics
FACULTY OF AGRICULTURE
www.ubkv.ac.in

From:
Dr. M.K. Debnath,
Chairman, Advisory Committee

Mobile: +91 8670937121
+91 8837481075
Email: mkanti1984@gmail.com

No..............................

Dated,..18/07/2023.....

Certificate

Certified that the thesis entitled **"FORECASTING YIELD OF MAJOR CROPS IN THE NORTHERN DISTRICTS OF WEST BENGAL"** submitted by **AJITH S** bearing Registration number A-2020-012-D towards partial fulfilment of the requirements for the award of Doctor of Philosophy (Agriculture) in Agricultural Statistics of the Uttar Banga Krishi Viswavidyalaya has been checked against plagiarism through URKUND software on **17th July, 2023** and that the similarity index has been achieved as **1%**, which is below the maximum tolerable range as per stipulation of this Viswavidyalaya. The thesis of **AJITH S** may be accepted for the award of the degree of Doctor of Philosophy (Agriculture) in Agricultural Statistics of the Uttar Banga Krishi Viswavidyalaya.

MKDebnath
18/07/2023

Place: UBKV, Pundibari
Date: 18/07/2023

(Dr. M. K. Debnath)
Chairman,
Advisory Committee

APPROVAL OF EXAMINERS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY (AGRICULTURE) IN AGRICULTURAL STATISTICS

We, the undersigned, having been satisfied with the performance of **Mr. AJITH S** (Reg. No.: A-2020-012-D) in the *viva-voce* examination on final evaluation of the Ph.D. (Ag.) thesis entitled **"FORECASTING THE YIELD OF MAJOR CROPS IN THE NORTHERN DISTRICTS OF WEST BENGAL"** conducted today, 15th March 2024 recommend that the thesis be accepted for the award of the degree of Doctor of Philosophy (Agriculture) in Agricultural Statistics.

Sl. No.	Name	Signature
1	Dr. Manoj Kanti Debnath **(Chairman, Advisory Committee)**	[signature] 15/03/2024
2	Prof. (Dr.) L. Hemochandra **(External Examiner)**	[signature] 15/3/24
3	Prof. (Dr.) Deb Sankar Gupta **(Member, Advisory Committee)**	[signature] 15/3/2024
4	Dr. Pradip Basak **(Member, Advisory Committee)**	
5	Prof. (Dr.) Subhendu Bandyopadhyay **(Member, Advisory Committee)**	[signature] 15/03/2024
6	Dr. Shyamal Kheroar **(Member, Advisory Committee)**	[signature] 15/03/2024
7	Prof. (Dr.) Deb Sankar Gupta **(Head, Department of Agril. Statistics)**	[signature] 15/3/24
8	I. **Dean, Post Graduate Studies**	[signature] 15/03/2024
9	Prof. (Dr.) D. Mukhopadhyay **(Dean, Facullty of Agriculture)**	[signature] 15.03.2024

Acknowledgements

This memorable occasion provides me a unique privilege to express my deepest sense of gratitude and indebtedness to the Chairman of my Advisory Committee **Dr. M.K. Debnath,** *Assistant Professor, Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya for his expert & intelligent guidance, scholarly advice, meticulous guidance, critical interaction and ever willing help throughout my doctoral degree program. His close monitoring and encouragement made me complete my work with perfection. I consider myself fortunate to have the privilege of being guided by a teacher who gave me complete freedom to express my thoughts and ideas not only in research work; but also, in every aspect. I admired his multi-faceted personality and ever-cool approach.*

I owe my sense of gratitude and unforgettable obligations to **Prof. (Dr.) D.S. Gupta**, *Professor and Head, Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya and Member of Advisory Committee for offering constructive criticisms, valuable suggestions, generous help & encouragement and providing necessary facilities. He is a pearl of a teacher and an asset to our department. I am indebted to him for his endurance, cordial behavior, salutary advice and incessant care towards me personally as well as in fulfilment of my goals.*

It gives me immense pleasure to record my sincere gratitude towards learned Members of the Advisory Committee, viz. **Dr. P. Basak**, *Assistant Professor, Department of Agricultural Statistics,* **Prof. (Dr.) S. Bandyopadhyay**, *Professor, Department of Agronomy and* **Dr. S. Kheroar**, *Assistant Professor, AINP on Jute and Allied Fibres, UBKV for their valuable suggestions, useful discussions and constructive counsel during the course of this investigation.*

I cordially offer my indebted remarks to other Teachers of the Department of Agricultural Statistics, UBKV namely, **Prof. (Dr.) S. Basak, Prof. (Dr.) A. Ghosh and Dr. S. Ojha** *who taught me various courses during my doctoral degree program.*

I express my deepest sense of gratitude to the Dean PG Studies **Prof. (Dr.) A.K. Choudhury** *as well as then Dean PG Studies* **Prof. (Dr.) P.K. Pal** *and to* **Prof. (Dr.) D. Mukhopadhyay**, *Dean, Faculty of Agriculture for rendering their support & help and provision of facilities during the course of my doctoral degree program.*

I am indebted to express my gratitude to **Prof. (Dr.) Ayon Roy**, *Deputy Registrar and all the staff of the office of the Deputy Registrar (Examination Section) and* **Mr. P.K. Paul**, *Assistant Librarian, Central Library, UBKV. It is worthwhile to mention* **Hasan da,** *cash section, for his fullest support and cooperation.*

*I extend my thanks to **Ratan Da** and **Ratna Di**, non-teaching staff of the Department of Agricultural Statistics for their timely action in all of my official proceedings during my doctoral program of study.*

*I am inexplicably indebted forever to my batchmates, **Chowa, Prem** and **Naffees** for their moral support, encouragement, fullest cooperation and ever-willing help during the course of my doctoral degree program. Blissful moments shared with them will remain cherishable.*

*I take this opportunity to express my thanks for the cooperation and help rendered by senior **Vaidheki** and Juniors, **Ragini, Vinay, Pavithra, Nishta, Neethu, Pooja, Karthik, Keerthana, Reshma, Amal, Anil, Sankhabrata, Shifat** and **Sumadeepika** as and when sought.*

*A special note of thanks to UBKV Tamil friends, **Pavithra**, **Vishnupriya** and **Velavan** as well as juniors **Kavya** and **Jayasri** for their joyful and memorable company. I extend my thanks to **Dinesh sir** and **Maharaja sir** for their help and guidance during the early phase of the Ph.D. program.*

*I find it hard to describe in words the incessant care and support that I am receiving from my friends **Dr. R. Karthik, Dr. M. Laxma Narayanan, A.C. Vigneshwar, G. Rajeswaran** and **K. Priyatharsini** in the last ten years and wish the same will last till the end of my life. They cheer me up in my difficult times and are with me in every moment of my happiness and sorrow.*

*In last but above all, my limited vocabulary seems inefficient in heralding the affection, great reverence and respect for my parents. The everlasting love, blessings, encouragement and motivation that I am receiving from my parents, Thiru. **C. Sakthivel** and Tmt. **S. Susila** as well as my sister and brother is beyond my capacity to express in words. All the good things in my life are because of their care and the way they have brought me up. They are the pivotal figures around whom my whole world revolves.*

I gratefully acknowledge the Government of West Bengal for the financial assistance provided to me in the form of Swami Vivekananda Merit-cum-Means Scholarship during the tenure of my doctoral program of study.

Any omission in this brief acknowledgment doesn't mean a lack of gratitude.

Place: UBKV, Pundibari

Date: 18/07/2023

S. Ajith 18/07/2023

Ajith S

CONTENTS

***

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

%	:	Per cent
°C	:	Degree Celsius
AIC	:	Akaike's Information Criterion
ANN	:	Artificial Neural Network
bales ha^{-1}	:	Bales per Hectare
DTR	:	Diurnal Temperature Range
et al	:	and other people
etc.	:	and so on; and other people / things
GDD	:	Growing Degree Days
i.e.	:	that is
ENET	:	Elastic Net
LASSO	:	Least Absolute Shrinkage and Selection Operator
LoS	:	Level of Significance
kg ha^{-1}	:	Kilogram per Hectare
MAE	:	Mean Absolute Error
MLR	:	Multiple Linear Regression
N/A	:	No value is available
nRMSE	:	normalized Root Mean Squared Error
OLS	:	Ordinary Least Square
PCA	:	Principal Component Analysis
PLSR	:	Partial Least Square Regression
R^2	:	Coefficient of Determination
RH	:	Relative Humidity
RBF	:	Radial Basis Function
RMSE	:	Root Mean Squared Error
RMSE-CV	:	Root Mean Square Error for Cross Validation
RR	:	Ridge Regression
RTD	:	Relative Temperature Disparity
SR	:	Stepwise Regression
SMW	:	Standard Meteorological Week
SSE	:	Sum of Squared Error
SVR	:	Support Vector Regression
t ha^{-1}	:	Tonnes per Hectare
Tanh	:	Tangent hyperbolic
Tmax	:	Maximum Temperature
Tmin	:	Minimum Temperature
viz.	:	Namely
WS	:	Windspeed

FORECASTING YIELD OF MAJOR CROPS IN THE NORTHERN DISTRICTS OF WEST BENGAL

ABSTRACT

The crop yield prediction gains a growing importance for all stakeholders in agriculture in order to make various agriculture related decisions. Crop yield prediction at regional level is essential for the formulation of location specific policies. In this present study, four northern districts of West Bengal namely, Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur are considered. Five majorly cultivated crops *viz.* Rice, Wheat, Potato, Jute and Rapeseed-Mustard are selected for the present study. Inter-annual crop yield variability largely depends on weather conditions that have been altered by climate change. The models based on weather variables provide a reliable forecast of crop yield. The extent of weather effect on crop yield is not only depends on magnitude of weather factors; but also, the distribution pattern of each weather factor over the crop season. Hence, it is required to give appropriate weightage to weekly weather conditions according to its effect on ultimate crop yield. In order to give weightage to the respective week's weather conditions, correlation based generated weighted indices have been used by many researchers in the past four decades. The path coefficient calculated from path analysis measures only the direct effect of independent variable on the ultimate response. Hence, path coefficient based weighted index is proposed in this study in a view of getting more precise and accurate weighted weather indices. Since, the crop yield is complexly associated with many environmental factors, different models with different functional forms have to be explored in order to obtain location specific best performing model. In this present study, the performance of Multiple Linear Regression (MLR), three penalized regression models *viz.* Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) and two machine learning models *viz.* Artificial Neural Network (ANN) and Support Vector Regression (SVR) have been evaluated. The inclusion of all the explanatory variables in the model makes the model to be complex and leads to overfitting issue generally. Hence, three variable selection or dimension reduction techniques namely, Stepwise Regression (SR), Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR) are employed in a view of avoiding overfitting issue, increasing speed of training algorithm and to avoid multicollinearity issue. The MLR, ANN and SVR models have been fitted under three variable selection conditions. Since the penalized regression models are doing inner variable selection by shrinking the coefficient values, there is no necessity of variable selection or dimension reduction. Twelve models are fitted for each crop in each district. The results revealed that there is a significant trend in the yield of all the five crops in all four districts except rapeseed-mustard yield in Cooch Behar. The significant increasing trend in crop yield can be attributed to the introduction high high-yielding varieties, improved cultivational practices *etc.* over the years. The proposed path coefficient based weighted indices are having predictable effect on crop yield along with the existing unweighted and correlation coefficient-based weighted indices. The number of variables selected from stepwise regression ranges between two to seven. Four to six principal components which explain more than eighty percent of the variability in the original variables are selected. The optimum number of PLSR components ranges from one to seven. The optimum number of hidden layer neurons in ANN is less than or equal to the number of input layer neurons. The optimum learning rate ranges between 0.02 and 0.10. The Tangent hyperbolic (Tanh) function is found to perform better as a hidden layer activation function in ANN. The hyperparameter-tuned SVR model using RBF kernel was found to perform better than linear and polynomial kernel. In comparison with OLS estimates, the regression coefficients of each index are shrunken in all three penalized regression models. In Ridge regression, the coefficient values are never shrunken to zero. But in LASSO and ENET models, some of the coefficients are shrunken to zero, thereby variable selection is accomplished. Only important variables are retained in the final LASSO and ENET models. The correlation and path coefficient-based weighted indices are retained in the final models. The penalized regression models provide stable performance in both training as well as validation stages for rice crop. For other crops, the machine learning models accompanied with PLSR components as input are found to perform well in both training and validation stages. The PLSR-SVR is found to perform better in most of the cases of wheat, potato, jute and rapeseed-mustard crops in four districts. The models using PLSR components as input performed better than PCA based models. If the dimension reduction is to fit the prediction model, the PLSR performs better than PCA. The best-fitted models have good forecasting ability with less percent deviation from actual yield. The yield of five major crops in four northern districts of West Bengal can be forecasted using the respective best-fitted models with the use of observed weather data.

Keywords: Yield, Weather Indices, Partial Least Square Regression, Statistical Models, Machine Learning Models, Forecasting.

18/07/2023

Dr. Manoj Kanti Debnath
Chairman, Advisory Committee

S. Ajith 18/07/2023

Ajith S
Reg. No.: A-2020-012-D

Chapter – I

INTRODUCTION

Chapter I

INTRODUCTION

The crop yield prediction gains growing importance for all stakeholders in agriculture starting from individual farmers to state, national and international level government as well as private organizations to make various agriculture-related decisions (Lobell *et al* 2006, Lecerf *et al* 2019, Vander-Velde *et al* 2019). The government requires advanced yield prediction of principal crops for policy-related decisions such as foodgrain availability, crop insurance, ensuring availability of farm inputs, raw material supply chain management to industries, price and market-related decisions *etc.* (Meinke and Stone, 2005, Kross *et al* 2020, Li *et al* 2022). The advanced yield prediction will be useful for the formulation of import and export policies to ensure an adequate supply of agricultural produce and to curb price hikes. At an individual farmer level, it is helpful to decide on what crop has to be grown, how much area to be allocated for each crop and other management as well as economic decisions (Basha *et al* 2020; Ansarifar *et al* 2021). Crop yield prediction at the regional level is essential for location-specific decision-making (Van Klompenburg, 2020, Rai *et al* 2022).

Crop yield over the years is affected by technological changes and weather variability. The technological factors such as the release of high-yielding varieties, improved cultivation practices *etc*, increase the yield smoothly over the years; but the weather variability both within and between seasons is the unmanageable source of variability in yield (Agarwal and Mehta, 2007). Inter-annual crop yield variability largely depends on weather conditions that have been altered by climate change (Delerce *et al* 2016). Climate change adversely affects agricultural production at the global level (Chandio *et al* 2022; Huang *et al* 2022). The changes associated with the effects of climate change and future population growth will be the major concern about food security (Hong *et al* 2019, Habib-ur-Rahman *et al* 2022). The pressure exerted on agricultural systems by the increasing climatic variability may constrain the attainment of food security (Aggarwal, 2008).

The average global temperature is expected to increase by 1.5°C over the next 20 years (IPCC, 2021). An increase in average temperature in the long run negatively affects crop yield (Jan *et al* 2021, Kumar *et al* 2021, Abbas, 2022, Ozdemir, 2022). The impact of climatic changes such as increasing temperature, diurnal temperature variation and interannual changes in the pattern and distribution of precipitation impacts the growth, photosynthesis and transpiration rates of crops which ultimately alter the yield (Tan *et al*

2021, Arun *et al* 2022). The most influencing climate anomalies on crop yield are temperature and precipitation-related parameters (Lacasa *et al* 2023).

In India, a significant temperature rise was observed over the years which is found to have an adverse effect on crop yield (Birthal *et al* 2014). There was a high magnitude of warming in winter and post-monsoon seasons (Khan *et al* 2017). 0.5°C -3.0°C increase in temperature leads to 0.6%-3.7% additional water requirement for the crops (Surendran *et al* 2021). There is a significant increasing trend in the temperature, a decreasing trend in the rainfall and profound seasonal variability in the weather phenomena over the last 30 years (Radhakrishnan *et al* 2017). Around 10% of districts of the country witnessed an increasing trend in rainfall and 8% of districts experienced a decreasing trend (Kaur *et al* 2017). As to the seasonal impacts, the winter season crops in India will be riskier (Mall *et al* 2006).

The impacts of climate change on crops vary by region and season and crop. The change in agroclimatic resources due to the effect of climate change significantly reduces the yield potential of different crops at global, regional and local scales (Hao *et al* 2018, Jia *et al* 2022). The changes in the climate-yield relationship are more pronounced at the local level than across relatively large regions (Trnka *et al* 2016). Often the reduction in agricultural production is not realized at the macro level, but has been a serious concern at the regional or district level (Mandal *et al* 2019, Bhardwaj *et al* 2022).

The Northern part of West Bengal state lies under the sub-Himalayan region. The districts of North Bengal are prone to vulnerable due to extreme climatic events during pre-monsoon and winter seasons (Ghosh *et al* 2017). It is projected that the average temperature in this region will increase at the rate of 0.027°C per year (Sam and Chakma, 2019). The changes in temperature trends are different in various seasons in this region. The cold and dry northern wind blows during winter, lowering humidity level substantially. An increasing trend in minimum temperature is greater than that of maximum temperature which causes larger diurnal temperature variation in this region (Datta and Das, 2019). There is an increasing trend in the diurnal temperature range in the winter season and a decreasing trend in the mean frequency of rainy days throughout the year except in the winter season in this region (Raha *et al* 2014).

The monsoon arrives in North Bengal earlier than South Bengal and the precipitation in North Bengal is more dispersed within a year in the sense that the participation is not only concentrated in the monsoon season (Chatterjee *et al* 2016). The dense forest cover and large plantations in this region act as a good source of evapotranspiration that results in well-distributed rainfall throughout the year (Kundu and Mondal, 2019). There has been a declining trend in the annual, monsoon, and post-monsoon rainfall and a non-significant increasing trend in the pre-monsoon rainfall in this region in the past century (Halder *et al* 2022). An increasing trend in winter temperature in this region impacts winter crop productivity in terms of higher water requirements due to increased evapotranspiration losses (Das *et al* 2019). Erratic weather and its undesirable impacts affect the agricultural sector by lowering the yield of principal crops in this region (Nag *et al* 2022).

Hence, it is necessary to couple meteorological information into the crop yield prediction mechanisms since the growth and development of crops are fully connected with weather factors (Semenov and Porter, 1995).

The extent of weather effect on crop yield depends not only on the magnitude of weather factors; but also, on the distribution pattern of each weather factor over the crop season (Laxmi and Kumar, 2011). The weather conditions during the critical period of crop development influence the yield significantly (Cardoso *et al* 2010). The same level of a particular weather parameter, say temperature affects crop growth in different ways in different crop stages (Asseng *et al* 2015, Chandarsiri *et al* 2022).

For example, for potential productivity of rice crop, the requirement of optimum temperature is 25.1°C-30.7°C during the germination stage, 27.2°C-29.6°C during the tillering stage, 22.4°C-31.0°C during panicle initiation and 25.0°C-27.6°C during anthesis (Sánchez *et al* 2014, Wang *et al* 2022). For mustard crop, the maximum temperature ranges between 20.0°C-27.3°C and minimum temperature ranging from 4.6°C- 11.8°C are required during the flowering stage and it is 11.3°C-20.4°C and 3.8°C-7.2°C during pod formation stage respectively (Kaur and Gill, 2017). Potato crop tuber induction is best at 15°C, tuber initiation is at 22°C and tuber setting is at 15°C (Struik, 2007). Heat stress during the flowering stage of wheat crop causes pollen and anther sterility that leads to undeveloped embryos which reduce the number of grains; heat stress during the grain filling stage reduces grain filling rate and ultimately affects grain yield (Dubey *et al* 2020). The movement of photosynthates to the sinks is adversely affected by heat stress during post-anthesis to seed filling stages of the crop that lowers the seed weight and seed yield (Kumar *et al* 2017).

Hence it is required to give appropriate weightage to weekly weather conditions according to its effect on ultimate crop yield. To give weightage to the respective week's weather conditions, Jain *et al* (1980) generated composite yearly weighted weather indices as weighted averages of weekly weather variables where weight is correlation-coefficient between weather variables in respective week and yield. Assigning appropriate weights in the composite index reduces the bias in regional-level climate-based models (Dhamija *et al* 2020). The correlation coefficient-based weighted weather indices can be successfully used as input variables to predict crop yield (Pandey *et al* 2016, Gupta *et al* 2018, Singh *et al* 2021).

The weather variables are not only correlated with the response variable; they correlate among themselves. The correlation coefficient between an independent variable and a response variable is the sum of the direct and indirect effects of the independent variable (Wright, 1923). Hence it is required to separate the direct influence of a particular variable on response (direct effect) from the indirect influence of that variable through other independent variables (indirect effect) on the ultimate response. Path analysis partitions the correlation coefficient into two components *viz.* direct and indirect effects (Li, 1956). The path coefficient calculated from path analysis measures only the direct effect of an independent variable on the ultimate response (Niles, 1923). The direct effect of an independent variable is part of its total effect that directly influences the response (Lleras, 2005). The difference between the path coefficient and correlation coefficient is that the latter is invariant of change of all other variables in the system whereas the former is not (Hope, 1971). Hence path coefficient-based weighted index is proposed in this study in a view of getting more precise and accurate weighted weather indices in combination with the existing correlation coefficient-based weighted indices.

The models based on weather variables provide a reliable forecast of crop yield. Many distinct climate-crop models are been used to estimate how the climate change process affects agricultural production (Rotter *et al* 2011). Statistical models are often used to understand complex associations between agricultural systems and different climatic parameters (Roberts *et al* 2017). These models are based on causal relationships between input and output variables (Siebert *et al* 2017). The empirically formulated statistical models with a smaller number of parameters could be used to predict crop yield using climate data (Lobell and Asseng, 2017, Kern, 2018, Gornott and Wechsung, 2022). Since crop yield is complexly associated with many environmental factors, different models with different functional forms have to be explored to obtain location-specific best-performing model (Mavromatis, 2016).

Linear regression is a standard statistical model to study the cause-and-effect relationship between one dependent variable and one or more independent variables. (Sagar and Cauvery, 2018). The penalized regression models such as Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) add penalty term in the Mean Square Error (MSE) function while estimating the parameters of the model. Due to the added penalty, the estimated parameter values are shrunken. Thus, the overall slope of the fitted model tends to decrease, thereby overfitting of the model can be avoided.

The penalized regression models are applied successfully to predict the yield of Rice (Das *et al* 2018), Wheat (Setiya *et al* 2022), Groundnut (Abhinaya *et al* 2021) and other crops.

Advanced machine learning methods provide insights into the pattern of crop response to climatic variations for practical crop yield prediction (Kang *et al* 2020). Machine learning models are promising methods to get optimized performance for regional-level crop yield prediction (Paudel *et al* 2020). The machine learning models observe the behavior of crops for the climate change effects to predict the yield (Hussain *et al* 2021). The machine learning regression algorithm is a better modeling approach for yield prediction using less data (Batool *et al* 2022).

Artificial Neural Network (ANN) is a widely used machine learning technique to capture non-linear associateship between a set of independent variables and a response variable (Vora *et al* 2014). The Back Propagation (BP) training algorithm is designed to reduce Mean Square Error (MSE) between the actual and desired output (Vora *et al* 2014). The activation function present in the hidden layer of the neural network should optimally update the weight of the BP algorithm in successive iteration to avoid convergence to any local minimum (Szandała, 2021). Hence, the selection of the appropriate hidden layer activation function is essential to model the complex association between weather factors and crop yield for a better predictive neural network model. The ANN model with one or two hidden layers is sufficient as well as efficient to predict crop yield using weather factors (Dahikar and Rode, 2014).

Vapnik (2000) developed a Support Vector Machine for classification purposes. Later the concept was modified for prediction context and the same was termed Support Vector Regression (SVR). The crucial role in fitting the SVR model is a selection of kernel function that captures nonlinear dynamics of closely related datasets (Qu and Zhang, 2016). The parameters or hyperparameters that are associated with the kernel function of the SVR model have to be optimized for better predictive models (Paidipati, 2021). To improve the performance of the SVR model for yield prediction using weather factors, the optimum kernel function along with their hyperparameters are to be identified.

While fitting a predictive model using many independent variables, variable selection plays an important role as it removes the variables that are redundant and do not influence the response variable significantly (Shahhosseini, 2020). This makes the algorithm work fast, reduces the problem of overfitting and allows to interpretation of the model easily (Kumar *et al* 2019). The number of variables can be reduced either by selecting the best subset of variables or transforming the original variable into derived variables in such a way that the few derived variables store most of the variation present in the original data (Vander-Maaten, 2009).

Stepwise regression can be employed to screen the weather variables that have significance on crop yield for fitting advanced yield prediction models (Han *et al* 2022). But in some situations, some variables that have a causal effect on the response will not be statistically significant and some variables become significant coincidently which leads to a good fit of the model for training data and poor performance with new data (Smith, 2018). Sometimes removal of some variables completely ignores the information that exists in those variables which reduces the predictability of the model (Olusegun, 2015). In most of the practical situations, the measured variables are prone to multicollinearity problems which leads to unstable estimates of parameters (Farrar and Glauber, 1967, Vatcheva *et al* 2016, Daoud, 2017).

To obtain orthogonal or uncorrelated independent variables, the Principal Component Analysis (PCA) is being employed traditionally as a dimension reduction technique (Bauer and Drabant, 2021). PCA forms a low-dimensional representation of high-dimensional data which is expected to describe as much variance in the original data as possible (Boneh and Mendieta, 1994). In PCA, the majority of the variability in the original variables is preserved in a few principal components (Pavithra and Vengadessan, 2020). PCA technique is used to replace a set of many measured variables with a smaller set of variables while fitting a predictive model (Jolliffe, 2003). The concept of replacing original explanatory variables with principal components in the statistical model was proposed by Hotelling (1957). The components that are associated with small variations can be safely dropped from regression analysis (Mosteller and Tukey, 1977)

The components that explain low variance are associated with very small latent root or eigenvalue. These components generally would not be considered for selection as this loss is not substantial. Hence, PCA can be adopted as a variable selection methodology such as backward elimination to eliminate the redundant component without decreasing the residual sum of squares of the models (Mansfield *et al* 1977). Since the principal components are orthogonal or uncorrelated to each other, the problem of multicollinearity can be mitigated using PCA (Carter *et al* 1977).

The Principal Component Analysis (PCA) considers only independent variables set while deriving components; it does not consider response variable. The principal components that explain less variance in the independent variable set are generally not selected for further analysis or modeling. But sometimes those components are important for explaining the response variable (Kung and Sharif, 1980, Jolliffe, 1982). In Partial Least

Square Regression (PLSR), the components are derived in such a way that it maximizes the covariance between a dependent variable and independent variable set (Firinguetti *et al* 2016). Hence, the PLSR components that are explaining majority variability in independent variable have the good predictability on response variable (Hu *et al* 2018). When the variable reduction is to fit a predictive model, PLSR performs efficiently than PCA (Yeniay and Göktaş, 2002, Maitra and Yan, 2008, Wang and Li, 2020).

In view of the above facts, the present study has been formulated to forecast the yield of major crops in some selected northern districts of West Bengal using weather indices with the following objectives,

1) To develop different weather and agrometeorological indices for major crops.
2) To select the indices using standard statistical procedures.
3) To fit and estimate the parameters of different statistical models for major crops.
4) To forecast the yield of major crops in the northern districts of West Bengal using respective best fitted models.

****

Chapter – II

REVIEW OF LITERATURE

Chapter II

REVIEW OF LITERATURE

A review of the literature provides guidelines from past studies which gives the foundation to a theoretical framework for the present investigation. The review of past literature allows the researcher to get an insight into the methods and procedures to be followed. With this aim, this chapter attempts a brief review of research works that were carried out in the area of crop-weather modeling and forecasting by various researchers at the national and international levels. The literature of related studies conducted hitherto has been grouped under the following three captioned categories for the convenience of better presentation and understanding of the subject matter.

2.1 Developing weather indices

2.2 Variable selection or dimension reduction methodologies

2.3 Crop yield forecasting models

2.1 Developing Weather Indices

Pioneering work in crop-weather modeling was done by Fisher (1924). The study revealed that the effect of change in rainfall on yield in successive weeks would not be an abrupt or erratic change, but it is an orderly one that follows some mathematical law.

Hendricks and Scholl (1943) modified Fisher’s technique by dividing the crop season into weekly intervals and assumed that a second-degree polynomial in the week number would be sufficiently flexible to express the relationship. The study concluded that the effect of weather parameters on different stages of the crop may help in understanding their response in terms of final yield and also provide a valid forecast of crop yield.

Jain *et al* (1980) modified Hendricks and Scholl's method by including a correlation coefficient between yield and weekly weather parameters as a weight for giving weightage to the respective week’s weather condition while developing generated weather variables.

Agarwal *et al* (1983) further included the joint effect of two weather variables in the model for forecasting rice yield. The year number is also included in the model to correct the long-term trend in the yield. The study revealed that the interaction of weather variables improves the predictive power of the crop-weather forecasting model.

Agarwal *et al* (1986) updated the Jain *et al* method by removing the trend effect from the yield while calculating the correlation coefficient as the trend effect on yield was significant. The study revealed that the model using generated variables obtained by yield adjusted for trend performs better than the unadjusted one. Also, the inclusion of the quadratic term of the correlation coefficient did not improve the performance of the model.

The above studies have provided a strong base for the crop-weather modeling approach. The different methodologies that had been developed for generating weather indices from the above studies are being applied by many researchers for crop yield forecasting using the generated weather indices as independent variables or input variables.

Bal *et al* (2004) compared two types of wheat yield forecasting models for Ludhiana district of Punjab state. The first model was fitted using only weather parameters as regressor variables and the second model was fitted by introducing the year number to capture technological changes along with the weather variables. The first model explained only 69 percent of the variation in the yield. However, by the introduction of technological trends in the model, the amount of variation explained by the model was increased to 87 percent. From the results, it was concluded that the weather variable along with the trend variable increases the predictive power of the yield prediction model.

Kumar and Bhar (2005) fitted a multiple linear regression model for predicting Indian mustard yield using weather parameters. The weather variables considered for the study were temperature, relative humidity, wind velocity and bright sunshine hours. The whole crop period was divided into critical phases. The weekly average of each weather factor was obtained for each phase. A composite weather variable was developed which is the weighted sum of weather variables of weeks of that phase where the weights is the correlation coefficients between yield and respective weather parameters of the week. Similarly, interaction terms were also developed as a weighted sum of products between two weather variables, where the weights being correlation coefficients of yield with product of weather variables in respective phases. Results of multiple linear regression on mustard yield with these developed composite weather variables indicated that the interaction terms had a significant effect. The interaction of maximum temperature and relative humidity was the most significant weather factor for forecasting the yield of Indian mustard.

Bazgeer *et al* (2007) fitted a wheat yield forecasting model using agrometeorological indices for the Hamedan district of Iran. The results from multiple linear regression revealed that daily minimum temperature, Growing Degree Days (GDD), accumulated difference of maximum and minimum temperature and sunshine hours were found to be the most significant factors that influenced the wheat yield. The above variables accounted 83 percent of the variation in the wheat yield.

Dhekale *et al* (2014) attempted to forecast rice yield in the Kolhapur district of Maharashtra. The unweighted indices and weighted indices were calculated where weight is the correlation coefficient between yearly detrended yield and weather parameters with respective weeks. Three generated variables *viz.* weighted interaction between morning relative humidity and minimum temperature, weighted interaction between morning relative humidity and maximum temperature and evening relative humidity only explained 89 percent of variation in rice yield.

Garde *et al* (2015) used weighted and unweighted weather indices along with their interactions for forecasting wheat yield in the Varanasi district of Uttar Pradesh. The weighted indices were calculated as weighted accumulations of the weather variables over weeks, where, the weight is the correlation coefficient between detrended crop yield and weather parameters in respective weeks. The weighted indices were found to be significant for forecasting wheat yield using Multiple Linear Regression.

Kumar *et al* (2016) applied a modified Hendrick and Scholl technique for forecasting paddy and sugarcane yield in south Gujarat. The results revealed that weighted weather indices and interaction indices had a significant effect on paddy yield. Unweighted did not have a significant effect on paddy yield in any district. Similar results were obtained for sugarcane crop also. The combined effect of weather parameters *viz.* minimum temperature and rainfall for paddy, minimum, maximum temperatures and rainfall for sugarcane were found to play a crucial role in determining the yield. Hence, it was concluded that the combination of weather variables and crop yield data is an appropriate option for yield forecasting.

Pandey *et al* (2016) attempted to forecast rice crop yield in the Faizabad district of Uttar Pradesh using five different weather indices approaches. The five different models were fitted as follows, unweighted indices, correlation coefficient-based weighted indices, unweighted interaction between two variables, weighted interaction between two variables and a combination of unweighted and weighted indices. Among all five fitted models, the model with unweighted interaction between two variables had a better predictive power with the highest R^2 (0.86) and lowest Root Mean Square Error (65.57) and predictive error (6.11%).

Banakara *et al* (2017) modeled kharif rice yield using weather parameters in the Valsad district of south Gujarat. The study revealed that the model with weather indices involving interaction between wind speed and rainfall, minimum temperature, maximum temperature and time trend performs better with the highest adjusted R^2 (0.61) and lowest Root Mean Square Error (381.45) values.

Jayakumar *et al* (2017) attempted to forecast coffee yield using weather indices in Kerala. The weather variables considered for the study were maximum temperature, minimum temperature, rainfall and relative humidity. Two types of climate indices were developed for each climate variable *i.e.* simple as well as weighted accumulation of climate variables where weights are correlation coefficients of climate variables in the respective week with yield. Similarly, interaction indices were also generated using products of weekly climate variables taking two at a time. From the results of stepwise regression, it was revealed that weighted indices of maximum temperature, weighted interaction indices of minimum temperature x relative humidity and maximum temperature x relative humidity were significantly affecting Arabica coffee yield which explains 78.3 percent of variation in yield. Similarly, weighted interaction indices of maximum temperature x minimum temperature and maximum temperature x rainfall significantly influenced Rustica coffee yield with a coefficient of determination of 0.529.

Poonam *et al* (2017) attempted to forecast wheat yield using weather parameters for the western agroclimatic zone of Haryana. Two types of models were fitted. To study the effect of individual weather variables, the yield was expressed as a quadratic function of the respective correlation coefficient of weather variables in different weeks with yield adjusted for trend. To study the joint effects of weather variables on wheat yield, interaction terms were also considered. The model using all seven weather parameters together was found to be the best-fitted model with the highest adjusted R^2 (0.86).

Gupta *et al* (2018) used twenty years of weather and yield data of wheat and mustard crops to fit a forecasting model for ten districts of western Uttar Pradesh. Simple and weighted weather indices were generated using weekly weather parameter values and their weighted values using correlation coefficients. The yield prediction model developed for all districts showed that minimum temperature was the most important parameter as simple and weighted indices in all the models for both crops. The coefficient of determination ranged between 0.30 to 0.81 for the models developed for the wheat crop and it ranged between 0.26 to 0.87 for the mustard crop. The percent error between observed and predicted yield was less than 10 kg ha^{-1} on validation of the fitted model for both the crops in all the districts implying that these models can be used for predicting the wheat and mustard yield using the respective model.

Rajavel *et al* (2018) applied a modified Hendricks and Scholl method by utilizing weather parameters from sowing time to a week before the harvest was employed to forecast rice yield in different districts of Chhattisgarh. The study revealed that rainfall, temperature, relative humidity and sunshine hours were the main weather parameters that affected the yield of rice in Chhattisgarh. Forecasting Error percent range from -10 to 6 in 2014 and -4 to 8 in 2015.

Sharma *et al* (2018) attempted to forecast wheat and soybean yield using generated weather variables in Malwa agroclimatic zone. Three types of variables were generated for each weather variable, one as a simple average of weather variables and the other one as a weighted average of weather variables with correlation coefficients of weather variables in respective weeks with yield as the weight. Similarly, for the joint effect of weather variables, interaction variables were generated as the products of weekly weather variables taking two at a time. The unweighted and weighted indices developed from rainfall, minimum temperature and maximum temperature were found to be the most influencing factors in almost all the districts. The model developed for the wheat crop at the pre-harvest stage had R^2 ranges from 0.38 to 0.73 at different locations; while it ranges from 0.33 to 0.75 for the soybean crop.

Das and Kumar (2019) developed a wheat yield forecasting model using weather variables in Banaskantha district of Gujarat. Simple and weighted weather indices were generated for each weather variable. The results from stepwise regression analysis revealed that weather parameters such as weighted relative humidity, weighted product of maximum and minimum temperature and unweighted product of bright sunshine hours and relative humidity were found to be the most influencing parameters for wheat yield. The fitted model had a significant R^2 (0.67) with a standard error of 168.40. There was a good agreement with actual and predicted yield while testing the performance of the fitted model with three years of independent data sets.

Chutiya *et al* (2021) used a modified Hendrick and Scholl model to forecast winter rice yield in fourteen districts in the Brahmaputra valley of Assam. The fitted models showed weighted and unweighted indices associated with maximum temperature, minimum temperature and interaction of maximum temperature and relative humidity increase the predictive power of the model for determining winter rice yield in most of the districts. The interaction of rainfall and maximum temperature showed a positive impact on crop yield in Lakhimpur and Kamrup districts. The technological trend was found to be a significant factor in the determination of rice yield in seven districts.

Singh *et al* (2021) attempted to forecast the yield of wheat, mustard and potato crops in western districts of Uttar Pradesh using unweighted and weighted weather indices. The results revealed that rainfall and maximum temperature were the most important parameters for determining the yield of wheat crop and maximum temperature was the most important parameter for mustard and potato yield. R^2 of the fitted models ranged between 0.44 to 0.96 for wheat crop, 0.57 to 0.87 for mustard crop and it ranged between 0.54 and 0.99 for potato. During the validation of the fitted models, the forecasted yield was in good agreement with the observed yield for wheat, mustard as well as potato crop.

2.2 Variable Selection or Dimension Reduction Methodologies

Keong and Keng (2012) established a stepwise multiple regression model by employing oil palm yield as the dependent variable and agrometeorological variables as independent variables. The study was conducted at Tun Razak Agricultural Research Centre at Tekam in the state of Pahang in Malaysia. The study revealed that palm age and Percent Average Water Holding Capacity (PAWHC) were significant at a 1 percent level of significance in stepwise multiple regression. The above variables cause 68 percent of the variability in the oil palm yield.

Yadav *et al* (2014) applied Principal Component Analysis for developing wheat yield foresting model using weather variables. Five weekly weather parameters were used to develop five unweighted and five weighted weather indices, ten unweighted and ten weighted interaction indices. Four different models were fitted using principal component analysis by utilizing the above weather indices. In the first model, all thirty indices were used in principal component analysis and the first six principal components accounted for 90.44 percent of total variance. For the second model, five weighted and five unweighted weather were used to extract the principal component from which the first three principal components explained over 75 percent of the total variance. In the third model, five unweighted weather indices and ten unweighted interactions were used and the first four principal components explained about 95.36 percent of the variance. In the fourth model, five weighted weather indices and ten weighted interactions used in principal component analysis and the first four principal components were accounted for about 85 percent of total variance. Four multiple linear regressions were fitted by taking yield as the dependent variable and the principal components selected from the above four methods were used as independent variables. The first model that was developed by utilizing all the thirty indices while extracting principal components was found to be the best-fitted model with the highest adjusted R^2 (0.80) and lowest RMSE of 3.378 and Percent Standard Error of Forecast (PSEF) of 3.19 percent.

Kumar *et al* (2014) fitted a stepwise multiple linear regression model to forecast the paddy and wheat yield of the Navsari and Bharuch districts of Gujarat. The generated weather variables were considered for the analysis. The results from stepwise regression analysis revealed that bright sunshine hours, evaporation and rainfall were significant at Navsari, while maximum relative humidity, rainfall and bright sunshine hours showed a significant effect on the paddy yield of Bharuch. Similarly, for wheat yield, only minimum temperature was significant for Navsari while maximum temperature and minimum temperature showed a significant effect for Bharuch.

Azfar *et al* (2015) applied Principal Component Analysis (PCA) to forecast mustard and rapeseed yield using weather parameters in the Faizabad district of Uttar Pradesh. Six different models were fitted as follows, using unweighted weather indices of six weather variables, using weighted weather indices of weather variables, using all 42 weather indices (including interaction indices) of weather variables, using weighted and unweighted weather indices, using unweighted and unweighted interaction of six weather variables and the final model with weighted and weighted interaction of weather variables to generate principal components. The result revealed that the model fitted with principal components generated using all 42 weather indices (including interaction indices) of weather variables was found to be the appropriate model with the highest R^2 (0.86) and lowest RMSE value (1.29 quintal per hectare). During validation, the same model has a very low percent deviation of 5.66 percent from the actual yield.

Goyal (2018) compared stepwise regression with weather parameters as regressors and with principal component scores as regressors to forecast wheat yield in selected districts of Haryana state. The result revealed that principal component analysis offered a substantial improvement over the least squares regression method in the presence of multicollinearity. The estimated yield from the selected models indicated a good agreement with actual wheat yield by showing only 2 to 10 percent average deviations in most of the districts except Panchkula district.

Hu *et al* (2018) used Partial Least Square Regression (PLSR) method to determine the factors controlling winter wheat yield. The experiment was done at Key State Agroecological Experimental Station in Shaanxi province of China. Cross-validation was used to determine the number of significant PLSR components. The results revealed that the first two components cumulatively explained about 94.3 percent of the total variance in wheat yield. The study suggested that the PLSR methodology is beneficial as it helps to eliminate co-dependency among independent variables and allows a less biased view of the contributions of factors to the wheat yield.

Banakara *et al* (2019) fitted forecasting models for kharif rice yield in the Navsari district of Gujarat using principal component analysis and multiple linear regression techniques. The principal components were extracted from fifteen unweighted and fifteen weighted weather indices. The decision on the number of principal components to be retained was taken by considering Kaiser's criteria and scree plot methods. Five principal components were retained as their eigenvalues were greater than unity. The cumulative variation of the first five components was about 93 percent. Multiple linear regression technique was applied by taking weather indices as predictors. A stepwise regression procedure was also applied to select significant indices. It was found that weighted interaction between rainfall and morning relative humidity, simple weighted morning relative humidity, weighted interaction between rainfall and maximum temperature and simple weighted minimum temperature along with time trend were found to have significant effect on kharif rice. From the results, it was found that multiple linear regression was superior as compared with principal component analysis for forecasting kharif rice yield.

Kumar *et al* (2019) attempted to compare stepwise regression with principal component regression for rice yield forecasting. From the results, it was found that principal component regression performs better with less mean square error and less percentage standard deviation than the stepwise multiple regression. The coefficient of determination of PCR was 0.92 percent which was higher than the coefficient of determination of stepwise multiple regression (0.85).

Sujaritha *et al* (2019) made a comparative study on Partial Least Squares Regression (PLSR) analysis and Principal Component Analysis (PCA) on a publicly available dataset named octane dataset, where spectral data of gasoline with 401 other attributes are provided. From the results, it was identified that the number of components required by the PLSR is comparatively lesser than PCA for the same level of accuracy of the model. The study concluded that the partial least squares regression method yields better prediction than the principal component regression model.

Bahrami *et al* (2020) used backward multiple linear regression to select the effective weather parameters influencing rainfed wheat yield in the Fars province of south Iran. The results indicated that the backward elimination procedure performs better than the forward selection in the multiple linear regression with a comparatively high R^2 value. The minimum relative humidity, rainy days and average relative humidity have a significant impact on rainfed wheat yield.

Duan *et al* (2020) applied Partial Least Squares Regression (PLSR) technique to identify factors affecting rice yield in Southern China. The study was conducted to quantify the influence of climatic, soil properties, geographic attributes and fertilizer types on rice yield and to identify the dominant control factors on rice yield in southern China. As these factors exhibit collinearity among themselves, PLSR was applied to elucidate the linkages between annual rice yield and 34 measured variables. These results indicate that these factors explained 57 to 85 percent of the variation in rice yield in different regions of Southern China. The study suggested that the PLSR method is beneficial as it partially eliminates the effect of multicollinearity among the independent variables and reduces bias in the contribution of the factors to rice yield.

Nain *et al* (2021) applied principal component regression to forecast rice yield using agrometeorological indices in the Karnal district of Haryana. The principal component analysis was used to address the problem of multicollinearity. Thirty indices composed of five unweighted weather indices, five weighted weather indices, ten weighted interaction indices and ten unweighted interaction indices were used to obtain principal components. The first twelve principal components that explain about 83 percent of the variation of the original variables were used in the final model. The multiple regression model with principal component scores as regressor showed a coefficient of determination of 0.79, whereas the multiple regression model with original variables was with low coefficient of determination (0.58). The study concluded that principal component scores derived from weather indices as predictor variables help to obtain better yield estimates.

2.3 Crop Yield Forecasting Models

Chen *et al* (2015) assessed the performance of Support Vector Regression (SVR) with Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) models for predicting rice yield in southwestern China. Three commonly used kernel functions of SVR *viz.* linear, polynomial and radial basis functions were explored. The radial basis function was the best-performing kernel in SVR with the lowest Mean Absolute Error (MAE) of 0.39 t ha^{-1}, Relative Root Mean Square Error (RRMSE) of 7.19 percent and the highest R^2 of 0.56. The results revealed that the Support Vector Regression (SVR) model with radial basis kernel outperformed ANN and MLR.

Das *et al* (2018) examined multiple linear and penalized regression models along with a neural network model for forecasting rice yield using weather parameters of the west coast of India. The study examined Stepwise Multiple Linear Regression (SMLR), penalized regression models *viz.* Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) and Artificial Neural Network (ANN) model solely and also in combination with principal components scores (PCA-ANN) to overcome the problem of multicollinearity. The tuning parameter alpha value was set at 1 for LASSO and it was 0.5 for ENET. The principal components having eigenvalues greater than one were selected which were able to explain more than 90 percent variability present in the original dataset in all the districts. The number of PCs retained ranged between 6 and 9. The optimum number of hidden neurons in ANN was varied between 1 and 12. It was identified that the number of hidden layers neurons was less in the PCA-ANN model as compared to the ANN model since the number of inputs was less in PCA-ANN. The results revealed that the ANN model performed better during calibration but it was worse during validation which indicated overfitting of the data. LASSO and ENET models outperformed the other models in both the calibration and validation stages. It was concluded that the better performance of the

LASSO and ENET model was due to penalization of the magnitude of regression coefficients with feature selection which prevents overfitting of the data. Hence, it was suggested that penalized models *viz.* LASSO and ENET can be effectively utilized for weather-based rice yield forecasting.

Oguntunde *et al* (2018) compared Support Vector Machine Regression (SVR) with Multiple Linear Regression (MLR) for the prediction of rice yield using thirty-three climatic variables of southern Nigeria. Both the models were fitted with original weather variables and principal components as well. The first eight principal components exhibited an eigenvalue greater than one which were cumulatively explained 83.10 percent variation in the original variables. In MLR, Solar radiation, humidity and evaporation were found to be the most significant variables which explained 68 percent of the variation in rice yield. While fitting MLR with principal component scores as independent variables, only the first principal component was significant which explained 64 percent variation. The SVR model using original variables yielded a coefficient of determination of 0.87, whereas the SVR model using principal component scores yielded a coefficient of determination of 0.75 only. The results revealed that SVR using all candidate predictor variables reduced the risk of over-fitting compared to MLR.

Parviz (2018) assessed the prediction accuracy of Multiple Linear Regression (MLR) and Support Vector Regression (SVR) models for barley crop yield in three provinces of Iran. Mean air temperature, minimum air temperature, maximum air temperature, precipitation and windspeed were used as input variables. Sensitivity analysis was done to select the optimal kernel for SVR. Linear, polynomial, Gaussian Radial basis and sigmoidal kernel function were explored. Low Relative Root Mean Square Error (RRMSE) and Mean Absolute Error (MAE) were obtained with linear kernel function in Zanjan and Gilan provinces and for Yazd provinces, minimum RRMSE and MAE were obtained with sigmoidal kernel function. The respective kernels were used to fit the SVR model. There was a decrease in RRMSE and MAE of SVR compared to MLR in all three provinces. The study concluded that SVR performed better for barley yield forecasting.

Alkaff *et al* (2019) explored the Support Vector Regression (SVR) model for forecasting wetland rice yield in South Kalimantan Province of Indonesia. Radial Basis Function (RBF) kernel with various combinations of parameters were examined. The results revealed that the SVR model fitted with RBF kernel with parameters of C=1.0, epsilon=0.002 and gamma=0.2 was the best fitted model with R^2 of 0.98 and RMSE value of 0.13. During validation of the model, the predicted yield was closer to the actual yield.

Hence, it was concluded that the Support Vector Regression model can be applied to forecast rice yield using climate variables.

Guo *et al* (2020) compared the performance of different machine learning models with traditional regression models for predicting rice yield using climate, phenology and geographic data of 75 Agrometeorological Experimental Stations in South China. Three machine learning methods namely Backpropagation Neural Network, Support Vector Machine Regression and Random Forest were fitted to the training dataset. It was found that the machine learning models performed better than the traditional regression model because of built-in non-linear relationships between rice yield and independent variables. More specifically, the performance of Support Vector Machine Regression and Random Forest models was better than Backpropagation Neural Networks. It was iterated that the poor performance of the backpropagation method may be due to the gradient steepest descent algorithm which would lead to a local minimum.

Kumar *et al* (2019) compared the Least Absolute Shrinkage and Selection Operator (LASSO) with stepwise regression for wheat yield forecasting. Thirty different unweighted and weighted indices that developed from the individual, as well as joint effects of five weather variables, were used as independent variables. Only two indices along with time trend were significant in stepwise regression and these variables were able to explain only 85 percent of variation in the wheat yield. The optimal LASSO model was fitted with five indices and a time trend which were able to explain 95 percent of the variation in the wheat yield. The percent error of the LASSO model was less than stepwise regression while validating the models. Hence, it was concluded that the LASSO model performed better than the stepwise regression model for forecasting wheat yield.

Singh *et al* (2019) applied the Least Absolute Shrinkage and Selection Operator (LASSO) model for forecasting wheat yield using weather indices. The performance of the LASSO model was compared with stepwise regression. The R^2 of LASSO regression was 0.85 which is slightly higher than stepwise regression (0.84). The Mean Square Error (MSE) and Root Mean Square Error (RMSE) of the LASSO model were 13385.20 and 115.70 respectively which were comparatively lower than the stepwise regression model. Hence, it was concluded that LASSO regression gives a better prediction model for wheat crop yield compared to stepwise regression.

Das *et al* (2020) evaluated different weather-based linear and nonlinear models for the prediction of coconut yield on the west coast of India. Linear models *viz.* Stepwise Multiple Linear Regression (SMLR), Principal Component Analysis in combination with SMLR (PCA-SMLR), Least Absolute Shrinkage and Selection Operator (LASSO) and

Elastic Net (ENET) and nonlinear models namely Artificial Neural Network (ANN) with original variables and in combination with principal component scores (PCA-ANN) were employed. The number of principal components retained in the PCA-SMLR model varied between four to seven which was able to explain about 90 percent variability of the original dataset. The number of hidden layers in ANN was five for Ratnagiri and North Goa and it was eleven for Dakshina Kannada and Kannur. The R^2 and RMSE of the fitted models ranged between 0.45-0.99 and 18-3624 nuts ha^{-1} respectively. The Absolute Percentage Error (APE) varied between 0.12 and 58.21 percent during validation. It was identified that the performance of SMLR and ANN using principal component scores as regressors was poor as compared with SMLR and ANN using original variables. This may be due to the exclusion of some components that explain less than 5 percent of variance which may have high predictive power in the regression model. The overall performance of the models was in the order of ENET > LASSO > ANN > SMLR > PCA-SMLR > PCA-ANN. It was concluded that the ENET model could be utilized for coconut yield prediction.

Abbas *et al* (2020) applied Linear Regression (LR), k-Nearest Neighbour (k-NN), Elastic net (ENET) and Support Vector Regression (SVR) for predicting potato tuber yield using soil and crop properties collected through proximal sensing. The study was conducted at six fields in Atlantic Canada. The whole dataset was partitioned into 80 percent and 20 percent for training and testing sets respectively. SVR model was fitted with a linear kernel function. For fitting ENET model, the penalty parameter value was fixed at 0.5. The highest R^2 was obtained in the SVR model with a slightly higher standard deviation of 0.07 in comparison with LR and EN. The lowest MAE and RMSE were also recorded in SVR model. Hence, it was concluded that the SVR model outperformed all other models. The reason behind the better performance of the SVR model was attributed to the better optimization provided by the SVR algorithm for a high number of variables. Also, additional functionality of the kernel in SVR improved the model's ability for predictions by understanding the nature of each feature.

Amaratunga *et al* (2020) compared the performance of three training algorithms in an Artificial Neural Network for predicting paddy yield in major growing areas of Sri Lanka using climate data. Three training algorithms considered for the study were Levenberg-Marquardt (LM), Scaled Conjugated Gradient (SCG) and Bayesian Regularization (BR). The results revealed that the LM training algorithm performed better compared to the other two algorithms. All three training algorithms were performed well with low Mean Square Error (MSE) in terms of validation. However, LM and SCG algorithms were better in terms of a smaller number of epochs and computational time. Hence, it was concluded that the LM training algorithm was performed better for determining the relationship between climatic factors and paddy yield with an MSE value closer to zero with a smaller number of epochs within less computational time.

Sridhara *et al* (2020) compared neural networks, stepwise linear regression and penalized regression models for forecasting rabi sorghum yield in Karnataka. Six statistical models *viz.* Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ENET), Principal Component Analysis in combination with Stepwise Multiple Linear Regression (PCA_SMLR), Artificial Neural Network (ANN) alone and in combination with PCA (PCA_ANN). The results revealed that LASSO and ENET models were found to be the best-fitted models with R^2 greater than 0.90 for most of the districts. During validation of the fitted model, LASSO was the best model based on the nRMSE value followed by ENET compared to PCA_SMLR, RR, ANN and PCA_ANN. It was attributed that LASSO and ENET models performed better by preventing overfitting of the model and reducing regression coefficient values by penalization imposed by them. It was concluded that LASSO and ENET models can be effectively utilized for district-level weather-based crop yield forecasting purposes.

Abhinaya *et al* (2021) applied penalized regression models to predict groundnut yield using weather indices in the Coimbatore district of Tamil Nadu. Penalized regression models *viz.* Ridge Regression, Least Absolute Selection and Shrinkage Operator (LASSO) and Elastic Net (ENET) were fitted by taking yearly groundnut yield as dependent variable and various weather indices as independent variables. While evaluating of performance of the fitted models using goodness of fit criteria, it was found that all three penalized regression models provided a better fit to the data. The multiple linear regression showed comparatively lesser R^2 and high error values due to lack of fit. Hence, it was concluded that the penalized regression models are the better alternatives to the classical linear regression model to deal with the problem of multicollinearity.

Guo *et al* (2021) made an attempt to predict rice yield using Artificial Neural Network (ANN) and Partial Least Square Regression (PLSR) based on agronomic and climatic variables in East China. Feed-Forward Backpropagation Neural networks (FFBN) was employed. The FFBN model having one hidden layer with 29 neurons showed better prediction performance. FFBN model had comparative higher R^2 of 0.86 than PLSR (0.75). The root Mean Square Error of PLSR and FFBN were 0.44 and 0.54 t ha^{-1} respectively. It was concluded that the nonlinear ANN model performed better than a linear model for yield prediction with a complex input context.

Paidipati *et al* (2021) explored various kernels in Support Vector Regression (SVR) to assess nonlinear patterns in rice yield. The SVR model was fitted with linear, polynomial and radial basis function kernels to predict rice yield data of India and also five leading rice-producing states. The performance of various kernels was assessed by cross-validation and hyperparameter optimization. A grid search optimization technique and 10-fold cross-validation were used to optimize the hyperparameters. For the overall India level, linear kernel with cost function C = 1.1 performed better with error validation such as RMSE of 27.52 and 31.05 and MAE of 23.05 and 22.72 for training and testing datasets respectively. For the training dataset of all the five states, the polynomial kernel performed better by allocating predefined parameters such as degree of polynomial $d\epsilon(1, 5)$, Cost $C\epsilon(0.05, 1.1)$ and scale parameter $\Upsilon\epsilon(0.05,0.5)$. Similarly, the polynomial kernel was the best with the same parameters for a testing dataset of four states *viz.* Uttar Pradesh, Punjab, Bihar and Tamil Nadu. For the testing set of West Bengal, the linear kernel performed better with cost function C=1.05. Hence, it was concluded linear and polynomial kernels can be exploited in the SVR model for forecasting rice yield of overall India and leasing states.

Shafiee *et al* (2021) compared Support Vector Regression (SVR) with the Least Absolute Shrinkage and Selection Operator (LASSO) model for predicting spring wheat yield. The study was conducted at Vollebekk Research Farm at the Norwegian University of Life Sciences in South-Eastern Norway. Separate models were fitted for 47, 54 and 70 days after sowing and also for pooled data. Radial Based Function (RBF) kernel function was found to be most appropriate in SVR for all dates, whereas the linear kernel function was chosen as the best kernel for pooled data. The LASSO model was able to predict the wheat grain yield with R^2 of 0.82, 0.81 and 0.86 and Mean Square Error (MSE) of 0.25, 0.23 and 0.19 respectively for 47, 54 and 70 days after sowing respectively. The SVR model was having R^2 of 0.80, 0.81 and 0.81 and MSE of 0.26, 0.26, and 0.23 respectively. The R^2 for pooled data was about 0.90 for both models and MSE values were also equal to 0.14. The above results revealed that the performance of both models was increased for pooled data. It was concluded that the LASSO model provides similar results as compared to SVR with much less computation time.

Vashisth (2021) compared three machine learning models namely, Artificial Neural Network (ANN), Support Vector Regression (SVR) and Random Forest (RF) for district-level mustard yield prediction using unweighted and weighted weather indices. The models were fitted using the variables selected from Stepwise Regression (SR) as well as dimension reduction by Principal Component Analysis (PCA). The results revealed that the models fitted using principal components as the input were performed better than the indices selected using stepwise regression. The SVR model performed well under both SR as well as PCA variable reduction conditions. Among the six fitted models, the PCA-SVR model outperformed with comparatively lesser RMSE, MAE and nRMSE in both the calibration and validation stages.

Wickramasinghe *et al* (2021) modeled rice yield with climate variables of major regions of Sri Lanka using various statistical and machine learning techniques. The fitted models were Artificial Neural Network (ANN), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Multiple Linear Regression (MLR), Robust Regression (RR) and Power Regression (PR). The performance of the fitted models was assessed in terms of coefficient of determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error Ratio (RSR) and bias value. ANN and GPR regression models performed better with R^2 values above 0.80 and comparatively less error value. Both the models were performed better while validation of fitted models with actual yield of all four stations. It was concluded that the machine learning techniques outperformed conventional statistical methods for modeling rice yield with weather variables.

Aravind *et al* (2022) made an attempt to forecast wheat yield using multiple linear regression, neural network and penalized regression techniques. Unweighted and weighted weather indices were used as independent variables and wheat crop yield was taken as the dependent variable for fitting the forecasting model. The models fitted were Stepwise Multiple Linear Regression (SMLR) with original variables and in combination with principal component scores, Artificial Neural Network (ANN) model alone and combination with principal component scores, Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET). Seventy percent of data were utilized for model training and the remaining data were used for model validation. On comparison of these fitted models, LASSO and Elastic Net performed better with an nRMSE value of less than 10 percent for four out of five locations. The study concluded that the penalized models *viz.* LASSO and Elastic Net performed better by avoiding the problem of overfitting as it reduces regression coefficients by imposing penalties.

Kakati *et al* (2022) made an attempt to forecast rapeseed and mustard yield using Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) techniques in fifteen districts in the Brahmaputra valley of Assam. Out of the twenty-seven years of data, the first twenty-five years of data were utilized for model calibration, and the remaining two years (2017‑18 and 2018‑2019) data were used for validation purposes. Thirty-one weather indices developed from five weather parameters and trend variables were used as input variables. The results revealed that ANN models were found to have greater R^2 values over MLR except in 3 districts. Similarly, a profound improvement was observed in the RMSE of ANN models. During validation of the fitted models, it was observed that the percent error between observed yield and predicted yield by ANN were within 5 percent in more than 80 percent of the districts. Hence, it was concluded that the weather-based ANN model was found to give a more accurate prediction of rapeseed and mustard yield in the Brahmaputra valley of Assam.

Kittichotsatsawat *et al* (2022) examined Artificial Neural Network (ANN) and Multiple Linear Regression (SMLR) models for predicting arabica coffee yield using weather parameters in the northern part of Thailand. The number of hidden layers in ANN was fixed as two. The optimum number of neurons in the hidden layers was determined by trial-and-error method to find the best ANN configuration. The ANN model with two hidden layers each with eight neurons was best performing configuration with the highest R^2 of 0.95 and the lowest Root Mean Square Error (RMSE) of 0.07. The MLR model was fitted by including all four weather parameters. The R^2 and RMSE were 0.92 and 0.01 respectively. The study confirmed the general phenomena of higher R^2 in the ANN model than that of the MLR model when large datasets were used. The study concluded that both MLR and ANN models were effective in forecasting arabica coffee yield.

Madeshwaran *et al* (2022) compared linear and non-linear models for predicting the yield of the West Coast Tall (WCT) cultivar of coconut in the Coimbatore district of Tamil Nadu using weather factors. The performance of three linear models *viz.* Ridge, Least Absolute Shrinkage Selection Operator (LASSO) and Elastic Net (ENET) regression methods were compared with the non-linear Artificial Neural Networks (ANN) model. The sigmoid activation function was utilized while fitting the ANN model. Results revealed that, in comparison with the nonlinear ANN model, linear models performed better. The coefficient of determination was comparatively high for ENET (0.98) followed by LASSO (0.96) and ANN (0.92). The RMSE, MAPE and MAE values were also found to be the least in the ENET model which were 58.98, 3.04 and 35.89 respectively. Hence it was recommended that the ENET model could be applied for the prediction of coconut yield.

Setiya *et al* (2022) compared five statistical and machine learning models for wheat yield forecasting for seven districts of Uttarakhand using weather indices. Five different models fitted were Stepwise Multiple Linear Regression (SMLR), Artificial Neural Network (ANN), Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ENET) and Ridge Regression (RR). During validation of the fitted models, the ANN model performed better as compared to other models with the lowest range of RMSE (0.01-0.47) and nRMSE (0.16-26.17) for all districts. The study also indicated that SMLR, LASSO, ENET and Ridge Regression models performed better during calibration but not good during validation of the model which may be attributed to the overfitting of the data. Hence, the study concluded that the ANN model was found to be the best model for forecasting wheat yield for the study region.

Tasan *et al* (2022) explored the performance of five machine-learning models to predict eggplant yield. The study was carried out at the Black Sea Agricultural Research Institute of northern Turkey. The machine learning models considered for the study were Artificial Neural Networks (ANN), Support Vector Machines (SVR), Random Forests (RF), k Nearest Neighbor (kNN) and Adaptive Boosting (AB). Ten vegetative indices were used as input variables. Principal components derived from these indices were also used as input. The performance of machine learning models under these two input conditions was examined. The first two principal components obtained from these ten indices could explain 88.68 percent of the variance in the original data. The activation function used in the ANN model were Tangent sigmoid and linear functions in the hidden and output layers respectively. While fitting the SVR model, polynomial and Radial Basis Function (RBF) kernel functions were used. The results revealed that all machine learning models performed better with higher accuracy using PCA-based inputs than original input variables. The best results were obtained with the PCA-ANN model with the highest coefficient of determination of 0.97, lowest Mean Absolute Error (MAE) of 274.81 kg ha^{-1} and Root Mean Square Error (RMSE) of 352.78 kg ha^{-1}. The result of the SVR model was also similar to the ANN model with slightly lower prediction performance. The RF model recorded a poor performance among the other fitted models.

Satpathi *et al* (2023) made a comparative study on statistical and machine learning models for forecasting rice yield in three districts of Chhattisgarh using unweighted and weighted weather indices. The performance of five models namely Stepwise Multiple Linear Regression (SMLR), Ridge regression, LASSO, Elastic Net and ANN were evaluated. The results revealed that the ANN and LASSO models performed better than other models in both the calibration and validation stages. The Ridge, SMLR and Elastic Net models were performed well during the calibration of the model but their performance was very poor during the validation of the models due to an overfitting issue.

Sridhara *et al* (2023) made an attempt to compare machine learning models such as Artificial Neural Networks (ANN), Support Vector Machines (SVR) and Random Forests (RF) with penalized regression models namely, Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) to predict Pigeon pea yield using weather parameters in eleven Pigeon pea growing districts of Karnataka. The study revealed that maximum temperature and relative humidity were identified as the most important weather factors influencing the yield of pigeon pea. Among the fitted models, ANN was selected as the best model for the prediction of pigeon pea yield with a good model fit and lower error values. The study concluded that the nonlinear machine learning models outperformed linear models which indicate the existence of a non-linear association between crop yield and weather parameters.

Vashisth and Goyal (2023) made an attempt to find the best machine-learning model to predict mustard yield in the major mustard-growing districts of Rajasthan using unweighted and weighted weather indices. Artificial Neural Networks (ANN), Support Vector Machines (SVR) and Random Forests (RF) models were fitted in combination with Stepwise Regression (SR) and Principal Component Analysis (PCA). Hence, six models were fitted for each district. The study revealed that the SVR model performed well under both variable selection methodologies followed by ANN. The study concluded that PCA-SVR and SR-SVR models can be used to predict the mustard yield at the region level with comparatively higher accuracy.

****

Chapter – III

MATERIALS AND METHODS

Chapter III

MATERIALS AND METHODS

This chapter aims to provide a description of the study area, crops under investigation, description of data and statistical tools and techniques employed in the analysis. The details of the materials and methods used in this study were discussed under the following headings.

3.1 Description of Study Area, Crop and Data.

3.2 Weather and Agrometeorological Indices.

3.3 Variable Selection or Dimension Reduction Methodologies.

3.4 Statistical and Machine Learning Models.

3.1 Description of Study Area, Crop and Data

3.1.1 Description of Study Area

The northern part of West Bengal lies under the Eastern Himalayan region. The North Bengal region lies within the latitude of 24°45'N to 27°20'N and longitude of 87°45'E to 89°50'E. The total geographical area of this region is 21859 square kilometers approximately which is 24.62% of the state's geographical area. The northern part of West Bengal comprises eight districts *viz.* Cooch Behar, Jalpaiguri, Malda, Uttar Dinajpur, Dakshin Dinajpur, Alipurduar, Darjeeling and Kalimpong. This region has a great inequality concerning topological and geographical features. Cooch Behar, Jalpaiguri and Alipurduar districts are characterized by a mixture of plains and hill landscapes while Darjeeling and Kalimpong districts are almost lies in hilly areas. Uttar Dinajpur, Dakshin Dinajpur and Malda districts are riverine plains.

This region is not well industrialized. Agriculture practices are the mainstay of the economy in this region (Paul and Paul, 2020). The plain areas of this region are being utilized for the intensive cultivation of various crops in different seasons. The overall soil status of this region is high organic matter, acidic pH, moderate nitrogen, phosphorus and potassium, high organic carbon and low sulphur and silicon content (Kunda *et al* 2021, Pati *et al* 2021).

The North Bengal region is dominated by cereal-based cropping system followed by the vegetable and jute-based system (Chatterjee *et al* 2013). There was a positive growth rate of 1.11%, 6.50% and 2.74% per year in the cultivation area of Wheat, Potato and Rapeseed-Mustard crops respectively in the North Bengal region in the last 40 years while Rice and Jute crops has experienced a negative growth in terms of cultivation area. The yield of Rice, Wheat, Potato, Jute and Rapeseed-Mustard achieved a positive growth rate of 2.72%, 1.04%, 4.53%, 2.56% and 1.15% per year respectively (Lepcha et al 2020). There is a great diversity in the productively of principal crops over different districts of this region (Aktar, 2015).

Based on the availability of long-term weather and crop yield data, four northern districts of West Bengal namely Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur are considered for the study. Cooch Behar and Jalpaiguri districts belong to the *Terai* zone and Malda and Uttar Dinajpur are in the Old Alluvial zone of West Bengal.

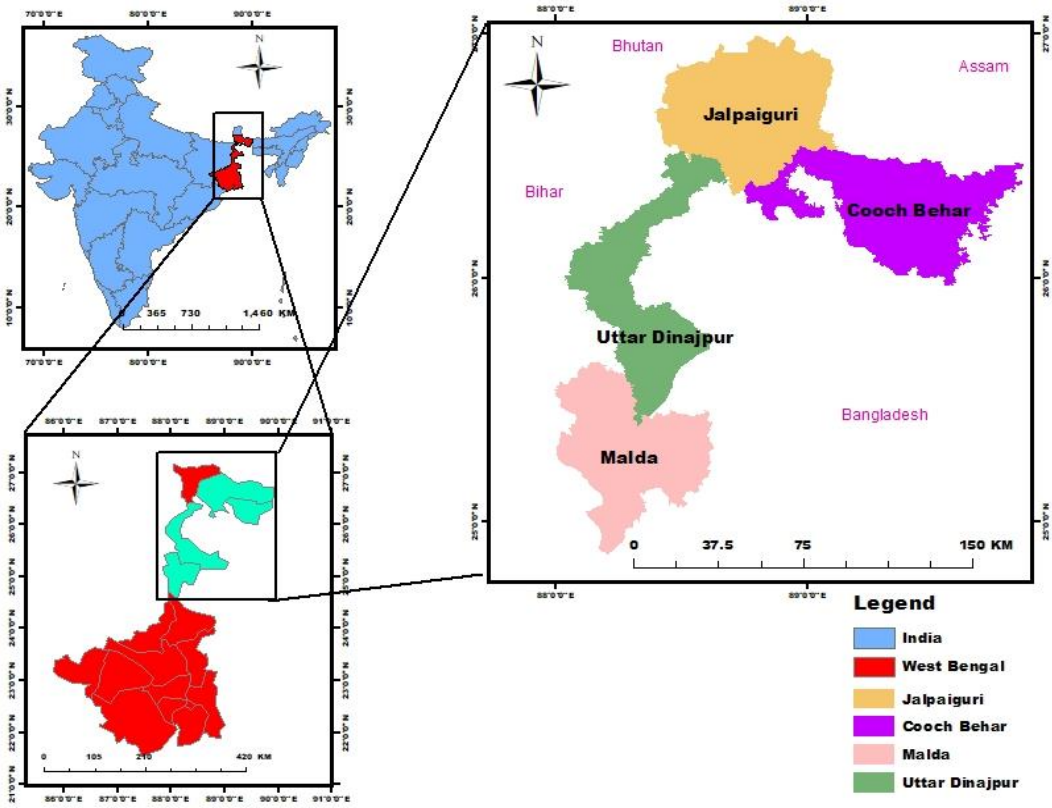

Fig 3.1 Study Area Map

3.1.2 Description of Selected Crops

Five majorly cultivated crops in terms of larger cultivation area in North Bengal *viz.* Rice, Wheat, Potato, Jute and Rapeseed-Mustard are selected for the present study.

3.1.2.1 Rice

Rice growing seasons in West Bengal are Aus (Autumn Rice), Aman (Winter Rice) and Boro (Summer). Winter rice is the major rice growing season in the Northern district of West Bengal. It is also known as Aman Rice. Generally, the sowing of winter rice starts during July to August and the ideal period for harvest is November or December. The name winter rice came as the harvesting takes place during the winter season. Aman season rice is the dominating rice season compared to the Aus and Boro seasons. The Aman Rice crop is grown in a 656.26 thousand hectares area in the north Bengal region that produces 16.96 lakh tonnes of rice grains with an average productivity of 2585 kg ha^{-1}. There was a declining trend in the cultivation area of Aman rice in this region because of crop diversification.

3.1.2.2 Wheat

Wheat crop is grown in the winter season as a rabi crop. Sowing of winter wheat takes place in the month of November-December and harvesting is done in March-April (Poddar *et al* 2022). The prevalence of cold weather conditions during the early stage of the crop and hot weather at the harvesting stage is suitable for wheat cultivation in this region. The wheat crop is grown in a 45.32 thousand hectares area in the north Bengal region that produces 1.09 lakh tonnes of wheat grains with an average productivity of 2403 kg ha^{-1}. There is a declining trend in wheat cultivation in the West Bengal state as a whole as well as in North Bengal. Due to the effects of environmental factors, a huge gap between the actual and potential yield of Wheat crop is observed in this region (Mukherjee and Huda, 2018).

3.1.2.3 Potato

Potato crop is mostly grown in the Rabi season in this region. Cooch Behar and Jalpaiguri are major potato-growing districts in North Bengal (Singh *et al* 2009). The potato crop is grown in 85.82 thousand hectares area in north Bengal region that produces 27.05 lakh tonnes of potato tubers with an average productivity of 31.52 tonnes ha^{-1}. The humid sub-tropical climate of North Bengal is conducive to the growth of potato crop (Roy, 2019). The environmental factors of the region greatly influence the growth and development of potato crop which plays a vital role in determining tuber yield. Drought Tolerance Index and Heat Tolerance Index are low in potato crop in this region (Pradel *et al* 2019). The number of favorable days for potato crop is expected to decrease due to the impact of climate change in this region (Luck *et al* 2012).

3.1.2.4 Jute

Jute is an important commercial cash crop grown by the majority of the marginal and small farmers of West Bengal state. West Bengal contributes four-fifths of jute production and three-fourths of cultivation area of the country (Barman and Anoop, 2023). The jute crop is grown in 134.88 thousand hectares area in the north Bengal region that produces 18.73 lakh bales of jute produce with an average productivity of 13.89 bales ha^{-1}. The crop is growing in the pre-kharif season from March to July in this region (Roy *et al* 2018, Kale *et al* 2023). Jute crop in West Bengal is often exposed to climate risk (Murthy *et al* 2022). Climatic condition during the growing period particularly temperature, rainfall and humidity plays a significant role in the occurrence major pests of Jute crop in this region (Rahman and Khan, 2012). The declining trend in rainfall during the March-April months necessitates high water requirements during the earlier stage of the crop for better establishment of the crop (Barman *et al* 2014).

3.1.2.5 Rapeseed-Mustard

Rapeseed-Mustard is one of the important oilseed crops that is grown in the Rabi season as a rainfed crop in North Bengal (Barma *et al* 2021). The rapeseed-mustard crop is grown in 141.81 thousand hectares area in the north Bengal region and produces 1.58 lakh tonnes of grains with an average productivity of 1112 kg ha^{-1}. The productivity efficiency of rapeseed-mustard crop is low in this region (Sur *et al* 2023). Temperature and humidity are the most influencing factors that cause the incidence of major pests and diseases of the crop. The major pests of the crop namely Sawflies and Aphids appear during the second fortnight of January and February in this region (Pal and Debnath, 2020).

3.1.3 Data Sources

The yearly yield data on selected crops in each of the selected districts from the year 1997-98 to 2020-21 have been collected from the Directorate of Economics and Statistics, Ministry of Agriculture & Farmers Welfare, Govt. of India. Weekly weather data of these districts have been collected from the Indian Regional Meteorological Centre, Kolkata. The weather parameters considered for the study are maximum temperature, minimum temperature, relative humidity, rainfall and windspeed.

Weekly weather data of the period in which a particular crop is grown are used to develop yearly weather indices for each crop. The crop growing periods of each crop in terms of Standard Meteorological Week (SMW) are given in table 3.1.

Table 3.1 The growing period of selected crops in the study region

S. No.	Crop	Crop growing period	Crop growing period in Standard Meteorological Week (SMW)
1.	Rice (Aman)	Second fortnight of June to first fortnight of November	26^{th} SMW to 45^{th} SMW
2.	Wheat	Second fortnight of November to first fortnight of March	47^{th} SMW of a year to 11^{th} SMW of next year
3.	Potato	Second fortnight of November to first fortnight of March	47^{th} SMW of a year to 11^{th} SMW of next year
4.	Jute	Second fortnight of March to second fortnight of June	12^{th} SMW to 25^{th} SMW
5.	Rapeseed-Mustard	Second fortnight of November to first fortnight of March	47^{th} SMW of a year to 11^{th} SMW of next year

3.2 Weather and Agrometeorological Indices

3.2.1 Calculation of Agrometeorological Indices to study the climate suitability of the study area for cultivation of crops

The agrometeorological indices are indicators for the climate suitability of a location for the cultivation of crops (Charalampopoulos, 2021). In the present study, the following agrometeorological indices have been calculated for each crop in each district to assess the climate suitability for potential cultivation of the selected crops in the study area in agrometeorological point of view.

3.2.1.1 Growing Degree Days (GDD)

A degree day is the mean temperature above the base temperature. It measures the abundance of thermal energy for crop growth. The GDD is an indicator of accumulated heat units in the crop. The unit of GDD is degree days (° days)

$$GDD = \sum \frac{T_{max}+T_{min}}{2} - T_b \qquad \text{...(3.1)}$$

where Tmax is the Maximum temperature, Tmin is the minimum temperature and T_b is the base temperature of a crop. The base temperature is the threshold temperature for crop growth below which the growth and development of the crop are affected significantly. The base temperature of the winter crop is 5°C (Verma *et al* 2003) whereas it is 10°C for Kharif crops (Chaudhari *et al* 2019).

3.2.1.2 Relative Temperature Disparity (RTD)

The Relative Temperature Disparity (RTD) is a relative measure of temperature variation which is expressed in percentage.

$$RTD = \left(\frac{T_{max}-T_{min}}{T_{max}}\right) \times 100 \qquad \text{...(3.2)}$$

3.2.1.3 Diurnal Temperature Range (DTR)

The Diurnal Temperature Range (DTR) measures the temperature variation within a day. The diurnal temperature difference is the measure of the thermal restriction factor for the potential cultivation of crops.

$$DTR = T_{max} - T_{min} \qquad \text{...(3.3)}$$

3.2.1.4 Number of Rainy Days

A day in which the rainfall is greater than or equal to 2.5 mm is termed as a rainy day as per the Indian Meteorological Department (IMD) (Kumar and Jain, 2011)

3.2.2 Calculation of different Weather Indices

Three types of yearly weather indices have been calculated for each crop in each district.

3.2.2.1 Unweighted-Indices

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

$$U_{ij} = \frac{\sum_{k=1}^{m} X_{ijk}}{m} \qquad \text{... (3.4)}$$

where, U_{ij} is unweighted weather index for j^{th} weather parameter in i^{th} year, X_{ijk} is the observation of j^{th} weather parameter in k^{th} week of i^{th} year and m is the number of weeks in which the crop is grown

3.2.2.2 Correlation-coefficient based Weighted Weather-Indices

The correlation coefficient based weighted indices are calculated as follows,

$$CC_{ij} = \frac{\sum_{k=1}^{m} r_{jk}.X_{ijk}}{\sum_{k=1}^{m} r_{jk}} \qquad \text{... (3.5)}$$

where, CC_{ij} is correlation coefficient based weighted weather index for j^{th} weather parameter in i^{th} year and r_{jk} is the correlation coefficient between crop yield and j^{th} weather parameter at k^{th} week.

The yield is to be detrended before calculating correlation coefficient if there exists a significant trend in the yield (Agarwal *et al* 1986). Detrended yield represents only actual effect of weather factors after removing trend causing factors over the years such as release of improved high yielding varieties and agronomic practices *etc*. (Huzsvai *et al* 2022).

Mann Kendal (MK) test is the most commonly used non parametric test to detect the presence of monotonic trend in the time series data (Wang *et al* 2020). The null hypothesis of the MK test is that there is no trend in the data (Gadedjisso-Tossou *et al* 2021). Modified Mann Kendal test is the robust test than MK test for trend estimation in the presence of autocorrelation in the data (Yue and Wang 2004, Hamed and Rao, 1997). The Sen's slope is another non-parametric method to estimate the magnitude of trend which measures the change per unit time (Kuriqi *et al* 2020).

The presence of trend in the crop yield data has been tested using Modified Mann Kendal test and the change per year in the yield has been quantified using Sen's slope method. The detrending of yield has been done by subtracting predicted yield from actual yield where the predicted yield has to be obtained by regressing the yield using year number as the explanatory variable (Sridhara *et al* 2020).

3.2.2.3 Path-coefficient based Weighted Weather-Indices

The correlation between an explanatory variable and response variable is the sum of direct and indirect effects of that particular explanatory variable. Path analysis partitions the correlation-coefficient into two components *viz.* direct and indirect effects (Freedman, 1987). The path-coefficient accounts for only the direct effect of an explanatory variable on the response variable (Alwin and Hauser, 1975). Hence path-coefficient based weighted-indices are proposed as the weighted average of weather variables where weight being the path-coefficient.

$$PC_{ij} = \frac{\sum_{k=1}^{m} p_{jk} . X_{ijk}}{\sum_{k=1}^{m} p_{jk}} \quad \text{... (3.6)}$$

where, PC_{ij} is path coefficient based weighted weather index for j^{th} weather parameter in i^{th} year and p_{jk} is the path-coefficient between detrended crop yield and j^{th} weather parameter at k^{th} week.

3.2.2.3 a) Path Analysis

Path analysis was first developed by Sewall Wright as a method to measure direct influence of causes as sperate path in a system to find degree to which variation of a given effect is determined by each particular cause. Path analysis is a multivariate statistical analysis widely applied in the context of cause-and-effect relationship in the system of interrelated variables (Deway and Lu, 1959).

In path analysis, independent variables or exogeneous variables are the determining variables and the variables determined by those are response variables or endogenous variables. The exogenous variables may be correlated among themselves. In contrast to the exogenous subset of variables, the total variation of the ultimate endogenous variable is assumed to be determined by some linear combination of exogeneous variables and other endogenous variables in the system if any, provided that the other endogenous variables in turn depend on the exogeneous variables and any further endogenous variables (Wright, 1960). That is, the ultimate endogenous variable is determined by the exogenous variables directly and/or through any other exogeneous variable(s). The undetermined variation of the ultimate exogeneous variable is accounted to the residual variable which is assumed to be uncorrelated with all the variables in the system.

Thus, path analysis examines an entire structure of interrelationship among a set of variables. It is able to handle not only simple direct causal relationships between variables but also any number of indirect causes and unanalysed relationships (Duffy *et al* 1981). Path analysis focuses on the problem of interpretation rather than a method for discovering causes (Wright, 1934).

The correlation between an independent variable and response variable is additive effect of direct and indirect effect of that particular independent variable. Further, the coefficient of correlation of two variables is the sum of products of path coefficients of all the chains that connecting the two variables (Niles, 1922, Li, 1968).

Indirect effects of independent variable on a response variable are those parts of its total effect which are transmitted or mediated via variables specified as intervening between cause and effect of interest in a model. Direct effect of is part of its total effect that it is not transmitted via any other intermittent variables.

Let there are 'k' independent variables X_1, $X_{2...}$, Xk and a response variable Y. The correlation between X_1 and Y is given as

$$r_{1Y} = \frac{Cov(X_1,Y)}{\sqrt{Var(X_1).Var(Y)}} = \frac{Cov(X_1,X_1+X_2+\cdots+X_k+R)}{\sqrt{Var(X_1).Var(Y)}} \quad \ldots(3.7)$$

$$= \frac{Cov(X_1,X_1)}{\sqrt{Var(X_1).Var(Y)}} + \cdots + \frac{Cov(X_1,X_k)}{\sqrt{Var(X_1).Var(Y)}} + \frac{Cov(X_1,R)}{\sqrt{Var(X_1).Var(Y)}} \quad \ldots(3.8)$$

$$= \frac{Var(X_1)}{\sqrt{Var(X_1).Var(Y)}} + \frac{r_{12}.\sqrt{Var(X_1).Var(X_2)}}{\sqrt{Var(X_1).Var(Y)}} + \cdots + \frac{r_{1k}.\sqrt{Var(X_1).Var(X_k)}}{\sqrt{Var(X_1).Var(Y)}} \quad \ldots(3.9)$$

$$= \frac{\sigma_1}{\sigma_Y} + r_{12}.\frac{\sigma_2}{\sigma_Y} + \cdots + r_{1k}.\frac{\sigma_k}{\sigma_Y} \quad \ldots(3.10)$$

$$r_{1Y} = p_{1Y} + r_{12}.p_{2Y} + \cdots + r_{1k}.p_{kY} \quad \ldots(3.11)$$

where,

p_{1Y}, p_{2Y},… and p_{kY} are the path coefficients of X_1, X_2… and X_k respectively.

r_{1Y}, r_{1Y},… and r_{1Y} are the correlation coefficient between Y and X_1, X_2… and X_k respectively.

R is unexplained part that is, Residual which is uncorrelated with any of the variables in the model.

Similarly,

$$r_{2Y} = r_{12}.p_{1Y} + p_{2Y} + \cdots + r_{2P}.p_{kY} \quad \ldots(3.12)$$

and so on,

$$r_{kY} = r_{1p}.p_{1Y} + r_{2p}.p_{2Y} + \cdots + p_{kY} \quad \ldots(3.13)$$

The solution of these system of simultaneous equations provides the direct and indirect effect of independent variables on response variable (Turner and Stevens, 1959). The above equations can also be written in the matrix form as

$$\begin{pmatrix} r_{1Y} \\ r_{2Y} \\ \vdots \\ r_{kY} \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ r_{k1} & r_{k2} & \cdots & r_{kk} \end{pmatrix} \times \begin{pmatrix} p_{1Y} \\ p_{2Y} \\ \vdots \\ p_{kY} \end{pmatrix} \quad \ldots(3.14)$$

Since r_{11}=r_{22} ... $= r_{kk} = 1$, the above equation can be rewritten as

$$\begin{pmatrix} r_{1Y} \\ r_{2Y} \\ \vdots \\ r_{kY} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{pmatrix} \times \begin{pmatrix} p_{1Y} \\ p_{2Y} \\ \vdots \\ p_{kY} \end{pmatrix} \quad \ldots(3.15)$$

C A P

Then, C = A x P …(3.16)

The matrix A is a symmetrical about the diagonal. The elements of column vector C represent estimated values of the correlation coefficients between dependent variables and response variable. The elements of the column vector P represent the path coefficients to be estimated as follows (Mitchell, 1969),

$P = A^{-1} x\ C$ …(3.17)

The diagonal elements in the resulted matrix P are the direct effects and the off-diagonal elements are the indirect effects.

Since there are five weather parameters, five unweighted, five correlation coefficients based and five path coefficient based weighted weather indices are developed as given in the table 3.2. The correlation coefficient based weighted indices are prefixed with CC and path coefficient based weighted indices are prefixed with PC.

Table 3.2 Description of Unweighted and Weighted Weather Indices

S. No.	Indices	Description
1.	**Tmax**	Unweighted index of Maximum Temperature
2.	**Tmin**	Unweighted index of Minimum Temperature
3.	**RH**	Unweighted index of Relative Humidity
4.	**WS**	Unweighted index of Windspeed
5.	**RF**	Unweighted index of Rainfall
6.	**CC_Tmax**	Correlation Coefficient based weighted index of Maximum Temperature
7.	**CC_Tmin**	Correlation Coefficient based weighted index of Minimum Temperature
8.	**CC_RH**	Correlation Coefficient based weighted index of Relative Humidity
9.	**CC_WS**	Correlation Coefficient based weighted index of Windspeed
10.	**CC_RF**	Correlation Coefficient based weighted index of Rainfall
11.	**PC_Tmax**	Path Coefficient based weighted index of Maximum Temperature
12.	**PC_Tmin**	Path Coefficient based weighted index of Minimum Temperature
13.	**PC_RH**	Path Coefficient based weighted index of Relative Humidity
14.	**PC_WS**	Path Coefficient based weighted index of Windspeed
15.	**PC_RF**	Path Coefficient based weighted index of Rainfall

3.3 Variable Selection or Dimension Reduction Methodologies

Inclusion of all the explanatory variables in the model makes the model to be complex and necessitate to estimate many parameter values. Inclusion of many variables prone to the multicollinearity problem that leads to unstable coefficients (Liu *et al* 2021). Hence, among the fifteen developed weather indices, the indices that have significant effect on yield are selected using three statistical methodologies namely, Stepwise Regression (SR), Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR).

3.3.1 Stepwise Regression (SR)

Stepwise regression is a classical variable selection methodology which is used to identify and select a useful subset of the important explanatory variables (Lewis-Beck, 1978, Johnsson, 1992). Stepwise regression selects explanatory variables based on their statistical significance. It is employed in stepwise manner on choosing the variables that give the best predictions by addition or deletion of variables at each step. Stepwise regression is the combination of both forward selection and backward elimination procedures (Ghani and Ahmad, 2010).

In forward selection, the regressor variables are added one by one successively in stepwise manner. The process begins with a null model. An explanatory variable that has highest partial correlation coefficient with the response variable is to be added in the first step. The significance of regression coefficient corresponding to the added variable is checked using F test. If the regression coefficient of added variable is significant and the addition of variable improves the overall predictability (R^2) of the model, then the variable is retained. Otherwise, the variable is discarded and another variable is to be added. In the second step, the explanatory variable that has second highest first order partial correlation coefficient given the first explanatory variable is added. The decision to retain the variable is taken based on significance of coefficient and improvement in overall R^2 after addition of second variable. In the third step, second order partial correlation coefficient given the first two explanatory variables is considered. This procedure is continued till all the variables are exhausted.

The backward elimination procedure starts with fitting a multiple linear regression model with all possible explanatory variables (full model). The least significant variables are to be discarded one by one in successive steps. The process continued until all the variables retained in the model are statistically significant.

The stepwise regression is the compromise between forward selection and backward elimination procedures in managing the limitations of both the methods (Huberty, 1989). In each step, every candidate variable that are already included will be revisited to check whether the inclusion of the variable in the model is still worth (Pope and Webster, 1972). It is possible to drop a variable that is found to be nonsignificant in the later stages given the significant variable is added (Henderson and Denison, 1989). In total, 2^k-1 models are fitted where k is the number of explanatory variables. The best subset model can be selected using statistical criteria such as highest adjusted R^2 or lowest AIC values (Ruengvirayudh and Brooks, 2016).

The F-test statistic calculated for the stepwise regression is given as,

$$F = \frac{MSR(X_i)}{MSE} \quad \ldots(3.18)$$

where, $MSR(X_i)$ is the Mean Square of the Regressor X_i and MSE is the Mean Square Error. The above F test statistic follows F distribution with (1, n-k) degrees of freedom at a specified level of significance, where n is the number of observation and k is the number of regressors. The candidate variable X_i is considered to be selected if the calculated F value exceeds the F calculated value. The variables selected using stepwise regression were used to fit the models.

3.3.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a multivariate statistical technique which is used in dimension reduction or variable reduction context (Wold *et al* 1987). The dimension reduction in PCA can be achieved by transforming a set of variables into a considerably smaller number of linear combinations of original variables (Lever *et al* 2017). The transformation is to be done in such a way that majority of the information present in the original set of variables is stored in few linear combinations (Dar, 2021). These linear combinations are termed as Principal Components (PCs). In PCA, the variances and covariance or correlation between the original set of variables are of interest. Hence the PCs are derived in such a way that most of the variations present in the original variable set is preserved in few PCs. Principal component has been applied in many fields for various purposes such as dimensional reduction, mitigation of multicollinearity problem, examining underlying variability existed in closely related datasets etc.

Let, X is a random vector that consists k random variables namely X_1, X_2, X_3,…X_k. It is assumed that the mean vector of X is 0 and variance-covariance matrix of X is real positive definite matrix $\sum$. The $\sum$ is a (kxk) matrix such that $(i,j)^{th}$ element is the covariance between i^{th} and j^{th} variables if $i \neq j$ and it is the variance of i^{th} variable if i=j. Let, $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the k non-zero eigenvalues of $\sum$ such a way that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$ and the corresponding eigenvectors are $\gamma_1, \gamma_2, \ldots, \gamma_k$.

For distinct λ_i's (i=1,2,…k), orthogonal matrix Γ of order (kxk) can be formed as

$$\Gamma = [\gamma_1, \gamma_2, \dots, \gamma_k] \quad \dots(3.19)$$

The matrix Γ diagonalizes ∑ such that

$$\sum = \Gamma\Lambda\Gamma' \quad \dots(3.20)$$

where, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots \lambda_k) = \Gamma'\Sigma\Gamma$. The orthogonal transformation of X vector to Z vector is represented as

$$Z = \Gamma' X \quad \dots(3.21)$$

where, $Z_1, Z_2, \dots, Z_k$ are the k components of Z which are termed as Principal Components (PCs). The rows of the matrix $\Gamma'$are eigenvectors which represent the orientation of principal components relative to original variables. The elements of each row or eigenvectors are the weights. These weights are termed as Loadings which describes the contribution of each variable to a particular principal component. A strong contribution of a variable to a particular component is represented by large loadings that may be either positive or negative. The sign of the loadings indicates whether a particular principal component is positively or negatively correlated with a variable. The diagonal elements of ∑ are the eigenvalues which is the variance explained by each component. The off-diagonal elements are zero as the principal components are uncorrelated with each other and covariance between them are zero.

Here, $$E(Z) = 0 \quad \dots(3.22)$$

$$V(Z) = E(ZZ') = E(\Gamma' XX' \Gamma) = \Gamma\Lambda\Gamma' = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots \lambda_k) \quad \dots(3.23)$$

The overall variability present in the original variable sets X can be taken as

$$tr(\Sigma) = tr(\Gamma\Lambda\Gamma') = tr(\Lambda) = \sum_{i=1}^{k} \lambda_i \quad \dots(3.24)$$

The variability stored in the transformed variables or principal components Z is

$$tr(\text{Z}) = \sum_{i=1}^{k} \lambda_i \quad \dots(3.25)$$

Hence the total variation stored in the principal components remain same as that of total variation existed in original set of variables X even after transformation.

Let, Z_i (i=1,2,…k) is the ith principal component of Z, then

$$E(Z_i) = 0 \quad \dots(3.26)$$

$$\text{V}(\text{Z}_\text{i}) = \lambda_\text{i} \quad \dots(3.27)$$

$$\text{Cov}(\text{Z}_\text{i}, \text{Z}_\text{j}) = 0 \text{ for all } i \neq j = 1,2, \dots k \quad \dots(3.28)$$

$$V(\lambda_1) \geq V(\lambda_2) \geq \cdots \geq V(\lambda_k) \geq 0 \quad \dots(3.29)$$

Since Σ is a positive definite matrix, λ_i is strictly positive.

Let, Z_1 is the first principal component corresponding to the eigenvalue λ_1 which is the linear combination of original sets of variables X_1, X_2, X_3,…X_k with elements of eigenvector γ_1 as weights. The Z_1 can be represented as follow,

$$Z_1 = \gamma_1' X = \alpha_{11}X_1 + \alpha_{12}X_2 + \cdots + \alpha_{1k}X_k = \sum_{i=1}^{k} \alpha_{1i}X_i \qquad \ldots(3.30)$$

where, $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1k}$ are the k elements of eigenvector γ_1. The vector $\gamma_1$is to be chosen such a way that $V(Z_1) = V(\gamma_1'X) = \gamma_1'\Sigma\gamma_1$is maximum. Here the constraint is the sum of squares of the elements of $\gamma_1$equals 1, that is, $\gamma_1'\gamma_1 = 1$. The standard approach to maximize $\gamma_1'\Sigma\gamma_1$ subject to the constraint $\gamma_1'\gamma_1 = 1$ is Lagrange multiplier technique. The quantity to be maximized here is

$$\gamma_1'\Sigma\gamma_1 - \lambda(\gamma_1'\gamma_1 - 1) \qquad \ldots(3.31)$$

where λ is a Lagrange multiplier. On differentiating the above equation with respect to γ_1 and equating to zero,

$$\Sigma\gamma_1 - \lambda\gamma_1 = 0 \qquad \ldots(3.32)$$

$$\Rightarrow (\Sigma - \lambda I_k)\gamma_1 = 0 \qquad \ldots(3.33)$$

where, I_k is the identity matrix of order k. λ is an eigenvalue of $\sum$ and the corresponding eigenvector is γ_1. In order to identify which among the k eigenvalues that give maximum variance of $\gamma_1'X$, the quantity to be maximized is

$$\gamma_1'\Sigma\gamma_1 = \gamma_1'\lambda\gamma_1 = \lambda\gamma_1'\gamma_1 = \lambda \qquad \ldots(3.34)$$

Hence, λ should be as maximum as possible. Thus, in order to get maximum variance of the first principal component, the eigenvector γ_1 corresponds to the largest eigenvalue of $\sum$ has to be chosen. Thus, $V(Z_1) = V(\gamma_1'X) = \gamma_1'\Sigma\gamma_1 = \lambda_1$is the largest eigenvalue of $\sum$.

The second principal component is derived as the weighted linear combination of original variables in such a way that the second component is uncorrelated with the first principal component. The eigenvalues are constrained to decrease monotonically from first component to last component. Hence, the second component has to be account for the maximum amount of remaining variation that already not been accounted for by first principal component. Thus, second principal component $Z_2 = \gamma_2'X$ that maximizes $V(\gamma_2'X) = \gamma_2'\Sigma\gamma_2$ subject to the condition $Cov(\gamma_1'X, \gamma_2'X) = 0$ has to be derived. The quantity to be maximized is

$$\gamma_2'\Sigma\gamma_2 - \lambda(\gamma_2'\gamma_2 - 1) - \phi\gamma_2'\gamma_1 \qquad \ldots(3.35)$$

where, λ and ϕ are Lagrange multipliers. On differentiating the above equation with respect to γ_2 and equating to zero results

$$\Sigma\gamma_2 - \lambda\gamma_2 - \phi\gamma_1 = 0 \qquad \ldots(3.36)$$

On multiplying γ_1 with above equation leads to

$$\gamma_1'\Sigma\gamma_2 - \lambda\gamma_1'\gamma_2 - \phi\gamma_1'\gamma_1 = 0 \quad \ldots(3.37)$$

Since first two terms in the above equation are zero and $\gamma_1'\gamma_1 = 1$, results $\phi = 0$. On substituting $\phi = 0$, the equation (3.35) can be rewritten as

$$\Sigma\gamma_2 - \lambda\gamma_2 = 0 \quad \ldots(3.38)$$

$$\Rightarrow (\Sigma - \lambda I_k)\gamma_2 = 0 \quad \ldots(3.39)$$

Again, λ is an eigenvalue of $\sum$ and the corresponding eigenvector is γ_2. Here the λ should not be equal to λ_1 since $\lambda=\lambda_1$ leads to $\gamma_2 = \gamma_1$, that violates the constraint $\gamma_1'\gamma_2 = 0$. Hence, the λ is the second largest eigenvalue (λ_2) of $\sum$ and the corresponding eigenvector is γ_2 that maximizes $V(\gamma_2'X) = \gamma_2'\Sigma\gamma_2$.

In similar fashion, the third, fourth,…,k^{th} principal component are derived as the weighted linear combination of original variables set X with eigenvalues λ_3, λ_4,…λ_k and their corresponding eigenvectors are $\gamma_3, \gamma_4, \ldots, \gamma_k$ respectively.

The percentage of variation explained by i^{th} principal component among total variation present in original dataset X is

$$\pi_i = \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i} \times 100 \quad \ldots(3.40)$$

The cumulative percentage of variation explained by first m principal components is (where $m \ll k$),

$$\pi_k = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{k} \lambda_i} \times 100 \quad \ldots(3.41)$$

The number of principal components derived is same as that of number of variables in original dataset. In general, the majority of variation in the original variable set X will be accounted for by first m principal components, where, $m \ll k$. Thus, few principal components store most of the variation present in many variables (Sclove, 2022). As the developed principal components are orthogonal or uncorrelated with each other, the PCA can be efficiently employed to avoid the multicollinearity problem (Eslamian *et al* 2010).

3.3.2.1 Determination of number of principal components to be selected

The key aspect of PCA is to select as few principal components as possible in such a way that the selected components store most of the variation present in the original variable set. The idea is to choose adequate number of components that represent the original variable set in optimal way. The decision on number of principal components to be selected for further analysis is taken based in the following criteria (Jolliffe and Cadima, 2016).

a) Proportion of total variance explained

The number of principal components that are cumulatively account for a predetermined proportion of total variation is to be selected (Rea and Rea, 2006, Cangelosi and Goriely, 2007, Wood *et al* 2021). The decision on proportion of total variation that is sufficient is subjective. The components that are explaining 80% variation cumulatively are selected for further modelling (Artigue and Smith, 2019).

b) Eigen value criteria (Kaiser, 1960)

The principal components that correspond to laten root or eigenvalue (λ) greater than one is to be selected (Olawuwo *et al* 2014). The idea behind is that the PCs which stores less than one variable worth information is redundant and the same has to be omitted in order to accomplish efficient dimension reduction.

c) Scree plot (Cattell, 1966)

In order to show the decreasing rate of variance that is explained by each additional principal components, the eigen values are plotted against their respective principal components in a graph which is termed as Scree plot. The number of principal components corresponding to a point before levelling-off of the curve (Elbow point) is to be selected.

If the PCA is employed in order to utilize the derived components as independent variables or input variables for fitting predictive models, the decision on number of components is to be oriented towards increasing predictability of the model. Selection of few components leads to poor model due inadequate representation original variables. On a contrary, selection of more components results overparameterized model and there is a question of dimension reduction (Valle *et al* 1999).

Principal Component Analysis is carried out using fifteen unweighted and weighted indices and a trend variable. Sixteen Principal Components (PCs) have been obtained as there are sixteen variables. The decision on the number of PC's to be retained was taken based on three criteria *viz.* proportion of variance explained, Kiser criteria and Scree Plot (Banakara *et al* 2019). Let, PC_1, PC_2, ..., PC_m be first m (m<16) principal components that explaining about 80 percent of the total variation in the original data and their eigen values are greater than one. Then these k principal components will be used as regressors or input variables in the models.

3.3.3 Partial Least Square Regression (PLSR)

PCA decomposes explanatory variables set (X) in order to derive components that best explain X. But there is no guarantee that the derived component explains the maximum variance in response variable (Y). The Partial Least Square Regression (PLSR) finds the components from the explanatory variables set (X) that best predict response variable (Y)

(Abdi, 2010). PLSR can be used as an alternative to PCA in the context of dimensionality reduction problem. The difference between PLSR and PCA is that the PLSR considers response variable (Y) also while deriving components from explanatory variables set (X) (Jia *et al* 2014). PCA components are extracted in such a way to maximize the variance of explanatory variables set while the PLSR components are extracted in such a way to maximize the covariance between response variable and explanatory variable sets.

The PLSR combines the features of Principal Component Analysis (PCA) and Multiple Linear Regression (MLR). Thus, the goal of PLSR is to predict a response variable or dependent variable using a set of independent variables. Here the prediction can be achieved by extracting a set of orthogonal factors from the independent variables set which is termed as PLSR components. PLSR is especially useful in the context of a greater number of explanatory variables and a smaller number of observations (Goktas and Akkus, 2020).

The PLSR searches a set of components that decompose both X and Y simultaneously with a constraint of maximizing the covariance between X and Y as maximum as possible (Abdi, 2003). It is followed by regression step in which the derived components are utilized to predict the response variable (Y).

The PLSR find a set of components (T) coupled with a specific set of loadings that decomposes X and simultaneously predict the response (Y). The double decomposition of independent variables set X and to predict the response variable Y is as follows,

$$X = TP' \text{ and } \hat{Y} = TBC' \qquad \ldots(3.42)$$

where, T is the PLSR score matrix such that $T'T = I$ and P is component loadings. B is a diagonal matrix that contains regression weight correspond to each independent variable in their diagonal elements. C is loading matrix of response variable. The columns of T are the PLSR components. These components are arranged in the order of variance of $\hat{Y}$ they are explaining.

The PLSR components are derived in using iterative application of Singular Value Decomposition (SVD). Each run of iteration provides a set of orthogonal components that decompose X and Y (Krishnan *et al* 2011).

Let, X and Y are normalized and mean cantered that are stored in the matrices X_0 and Y_0 respectively. The matrix of covariance or correlation between X_0 and Y_0 are calculated and stored in R_1.

The SVD is applied on the R_1 in order to decompose it into two sets of orthogonal singular vectors W_1 and C_1 with a corresponding singular value Δ_1 which is represented as below,

$$R_1 = W_1 \Delta C_1{}' \quad \ldots(3.43)$$

The first columns of W_1 and C_1 are taken as the first pair of singular vectors which are represented by w_1 and c_1 respectively. The corresponding singular value is represented by δ_1 which is the first diagonal elements of the matrix Δ. The δ_1 represents the maximum covariance between these two singular vectors w_1 and c_1.

First PLSR components is derived as below,

$$t_1 = X_0 w_1 \text{ such that } t_1{}' t_1 = 1 \quad \ldots(3.44)$$

The loadings of X_0 on the first component t_1 is as follows,

$$p_1 = X_0{}' t_1 \quad \ldots(3.45)$$

The estimate of X from first PLSR component is represented as follows

$$\hat{X}_1 = t_1{}' p_1 \quad \ldots(3.46)$$

Similarly, the decomposition of Y is represented as follows,

$$u_1 = Y_0 c_1 \text{ and } \hat{Y}_1 = u_1{}' c_1 \quad \ldots(3.47)$$

Since the first component t_1 is derived in such a way to the maximum covariance between two singular vectors w_1 and c_1 that are stored in covariance matric (R_1) of X and Y, the first PLSR component explains as maximum variation as in both X and Y.

In the similar fashion, second component t_2 is derived such that it is uncorrelated with t_1. The process continues up to k components, where k is the number of independent variables. In general i^{th} is represented as,

$$t_i = X w_I \quad \ldots(3.48)$$

such that Cov(t_1,Y)=maximum, $t_i{}' t_i = 1$ and $t_i{}' t_j = 0$ for $i \neq j$.

As the PLSR components are derived such that first component is expected to explain maximal variation in both X and Y. The amount variation explain by the PLSR components decreases as the component are increasing (Cheng and Wu, 2006). Hence variation explain by the later variable are generally low and inclusion of such components would not increase the predictive power of the model as well as it leads to overfitting of the model.

A cross-validation procedure is commonly employed in PLSR for determining the appropriate number of components (Han *et al* 2002). The optimal number of PLSR components corresponded to a minimal RMSE-CV (Krol, 2017, Mehmood *et al* 2020).

The PLSR yields maximum reduction in number of variables (Hoskuldsson, 1988). The higher predictive power compared to PCA is the most desirable property of PLSR analysis (Wakeling and Morris, 1993). Since the PLSR components derived are orthogonal to each other, the PLSR can handle the highly collinear data (Shetty and Gislum, 2011). The PLSR is a robust technique when the independent variables are highly collinear and the number of variables is greater than number of observations (Yang *et al* 2006, Bhat and Vidya, 2018, Cook and Forzani, 2020, Zifarelli *et al* 2020).

3.4 Statistical and Machine Learning Models

3.4.1 Multiple Linear Regression

Let, there are k independent variables X_1, X_2, X_3,…X_k and Y is the response or dependent variable. The linear regression is given as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + \varepsilon \qquad \ldots(3.49)$$

or

$$\mathrm{Y} = \beta_0 + \sum_{j=1}^{k} \beta_j X_j + \varepsilon \qquad \ldots(3.50)$$

where, β_1, β_2, β_3,…β_k, are the unknown parameters corresponds to X_1, X_2, X_3,…X_k respectively. The coefficient β_j represents the effect size of j^{th} independent variable on the response. That is, for a unit change in the j^{th} independent variable (while keeping other variable fixed), the observed change in the response will be β_j. The ε is the unknown term or error term which represents the portion of response that is not explained by the functional part $\sum_{j=1}^{k} \beta_j X_j$ of the model.

If there are n observations on Y and X, then the equation can be written as

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_p X_{ik} + \varepsilon_i \quad \text{for i=1,2,…n} \qquad \ldots(3.51)$$

Alternatively, it can be written in matrix form as,

$$Y = X\beta + \epsilon \qquad \ldots(3.52)$$

where,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \qquad \mathrm{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}; \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \text{ and } \epsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

From the model assumption, we have

$$E(\epsilon)=0 \text{ and } Cov(\epsilon)=\sigma^2 I_n \epsilon \qquad \ldots(3.53)$$

where, I_n is the identity matrix of the order n.

In the equation, the unknown parameter vector β has to be estimated. The Ordinary Least Square (OLS) method is the most commonly used method of estimation. The OLS estimator is consistent and optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated and there is no multicollinearity. Under these conditions, the method of OLS provides minimum-variance mean-unbiased estimation when the errors have finite variances. Under the additional assumption that the errors be normally distributed, OLS is the maximum likelihood estimator.

The object is to find a vector $\beta' = (\beta_1, \beta_2, \dots \beta_p)$ such that it minimizes the sum of square of residuals.

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) \quad \dots(3.54)$$

The minimum is always exists as S(β) is real valued, convex and differential function. The above equation can be rewritten as

$$S(\beta) = YY' + \beta'X'X\beta - 2\beta'X'Y \quad \dots(3.55)$$

Differentiate S(β) with respect to β

$$\Rightarrow \frac{\partial S(\beta)}{\partial \beta} = 2X'X\beta - 2X'Y \quad \dots(3.56)$$

$$\Rightarrow \frac{\partial^2 S(\beta)}{\partial \beta^2} = 2X'X > 0 \quad \dots(3.57)$$

which is a non-negative. Hence the sum of square of residual S(β) function has a minimum and the minimum is obtained by equating the first order derivative to zero.

$$\frac{\partial S(\beta)}{\partial \beta} = 0 \quad \dots(3.58)$$

$$X'X\beta = X'Y \quad \dots(3.59)$$

which is the normal equation. Since it is assumed that rank (X)= k (full rank), then $X'X$ is positive definite. The unique solution of the normal equation is

$$\hat{\beta} = (X'X)^{-1}X'Y \quad \dots(3.60)$$

which is the Ordinary Least Square (OLS) estimator of β. Since $\frac{\partial^2 S(\beta)}{\partial \beta^2}$ is at least non negative definite, $\hat{\beta}$ minimize S(β).

The least square estimators are unbiased estimators of the parameter.

i.e.

$$E(\hat{\beta}) = E[(X'X)^{-1}X'Y] \quad \dots(3.61)$$

$$= (X'X)^{-1}X'E(Y)$$

$$= (X'X)^{-1}X'E(X\beta + \epsilon)$$

$$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'E(\epsilon)$$

$$E(\hat{\beta}) = \beta \;\; [\therefore E(\varepsilon) = 0 \; and \; (X'X)^{-1}X'X = I] \quad \ldots(3.62)$$

Hence, the OLS estimators are unbiased estimators of β. The variance of the OLS estimator is derived as,

$$V(\hat{\beta}) = V[(X'X)^{-1}X'Y] \quad \ldots(3.63)$$

$$= (X'X)^{-1}X'V(Y)[(X'X)^{-1}X']'$$

$$= (X'X)^{-1}X'V(Y)X[(X'X)^{-1}]'$$

$$= \sigma^2(X'X)^{-1}X'X[(X'X)^{-1}]'$$

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad \ldots(3.64)$$

Hence, the OLS estimators are unbiased estimators of parameters with variance $\sigma^2(X'X)^{-1}$

3.4.1.1 Problems in OLS

The OLS estimators were unbiased estimators. But its variance is generally very high. The existence of slight to moderate correlation between independent variables leads to the problem of multicollinearity and the corresponding regression coefficient were unstable due to high SE. If there exist a strong multicollinearity between independent variables, the regression coefficient cannot be estimated (Dorugade and Kashid, 2010).

One of the common problems arises with OLS method is the tendency to overfit the data when there is too much noise caused by correlated variables. Overfitted model describe the data very well. But its performance for the new data is generally poor (Zhang, 2014). The penalized or regularized regression models are used to mitigate these problems.

3.4.2 Penalized Regression models

This is also known as regularization. These methods involve the introduction of an additional information in the form of penalty in addition to the Residual Sum of Squares (RSS) function while estimating parameters. The general form of the function that has to be minimized while estimating the parameters of the penalized regression is as given below,

$$\sum_{i=1}^{n}\left(y_i - \sum x_{ij}\beta_j\right)^2 + \lambda * Penalty(\beta) \quad \ldots(3.65)$$

The first part of the above equation is the RSS function of the OLS estimation. In addition to the RSS of OLS, the penalty term also to be minimized while estimating the parameters. The penalty term differs based on the penalized regression models. The term λ is the penalty parameter or regularization parameter which governs the amount of penalty to be imposed. If λ=0, then the entire penalty term become zero, which is equal to OLS estimator. If the λ is low, it adds less penalty and the amount of penalty increases as λ increases.

The idea of the penalized regression model is to scarifies the unbiasedness property in order to reduce the variance of the estimator. That is, in order to decrease the variance of the estimator, the bias of the estimator is increased for a small quantity deliberately. Due to the addition of penalty some of the coefficients are shrunken. Thus, the overall slope of the fitted line decline (Wang *et al* 2022). This regularization improves the performance of the linear regression models.

The performance of three penalized regression models *viz.* Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) are examined.

3.4.2.1 Ridge Regression (RR)

Ridge regression imposes L_2 penalty on linear regression. The L_2 penalty is sum of squared coefficients. Then the function that has to minimized is

$$S(\beta)_{RR} = \sum_{i=1}^{n}\left(y_i - \sum x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{k}\left(\beta_j\right)^2 \quad \ldots(3.66)$$

This function can be written in matrix notation as below,

$$S(\beta)_{RR} = (Y - X\beta)'(Y - X\beta) + \lambda\beta'\beta \quad \ldots(3.67)$$

$$= (Y' - \beta'X')(Y - X\beta) + \lambda\beta'\beta$$

$$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta + \lambda\beta'\beta$$

$$S(\beta)_{RR} = Y'Y - 2Y'X\beta + \beta'X'X\beta + \lambda\beta'\beta \quad \ldots(3.68)$$

On differentiating the above function with respect to β

$$\Rightarrow \frac{\partial S(\beta)_{RR}}{\partial \beta} = -2Y'X + 2X'X\beta + 2\lambda\beta \quad \ldots(3.69)$$

$$\Rightarrow \frac{\partial^2 S(\beta)_{RR}}{\partial \beta^2} = 2X'X + 2\lambda I = 2(X'X + 2\lambda I) > 0 \quad \ldots(3.70)$$

Hence function $S(\beta)_{RR}$ has a minimum. On equating the first derivative to zero

$$\frac{\partial S(\beta)_{RR}}{\partial \beta} = 0 \quad \ldots(3.71)$$

$$\Rightarrow -2Y'X + 2X'X\beta + 2\lambda\beta = 0$$

$$\Rightarrow \hat{\beta}_{RR} = (X'X + \lambda I)^{-1}X'y \quad \ldots(3.72)$$

Even if X is not a full rank matrix, $X'X + \lambda I$ is positive definite. This is very useful when k>n. In this case, $X'X$ can be at most rank of n<k. Thus $X'X$ is not invertible. Eventually, the least square estimate is undefined. However, $X'X + \lambda I$ is always invertible

and the ridge estimate always exists. This is an advantage of ridge regression over OLS. For λ>0, the RSS of the ridge regression is less than that of OLS.

The penalty parameter shrinks the coefficients. If $\lambda \to \infty$ the coefficients shrink towards zero. If the λ is close to zero, the estimate is close to OLS estimate. The penalty parameters λ is ranged between zero and infinity. However, the coefficient values will not be shrunken to zero unless $\lambda = \infty$. Hence all the variables will be retained in the final model. Thus, variable selection will not be accomplished in ridge regression. The value of λ for which the reduction in the variance is not exceeded the squared bias is the optimum value of the penalty parameter. The optimum choice of penalty parameter value is determined by cross validation procedure.

Another advantage of ridge regression is bias-variance tradeoff. As λ increases, bias increases and variance decreases.

$$E(\hat{\beta}_{RR}) = E[(X'X + \lambda I)^{-1}X'Y] \qquad \ldots(3.73)$$

$$= (X'X + \lambda I)^{-1}X'E(Y)$$

$$E(\hat{\beta}_{RR} = (X'X + \lambda I)^{-1}X'X\beta$$

$$E(\hat{\beta}_{RR}) \neq \beta \qquad \ldots(3.74)$$

Hence, the ridge estimator is not an unbiased estimator. The relationship between Ridge and OLS estimator can be derived as below,

$$\hat{\beta}_{RR} = (X'X + \lambda I)^{-1}X'Y \qquad \ldots(3.75)$$

$$= (I + \lambda I)^{-1}X'Y$$

$$= (1 + \lambda)^{-1}IX'Y$$

$$= (1 + \lambda)^{-1}(X'X)^{-1}X'Y$$

$$\hat{\beta}_{RR} = (1 + \lambda)^{-1}\hat{\beta} \qquad \ldots(3.76)$$

Thus, the ridge estimator is inversely proportional to the amount of penalty. The absolute value of ridge estimate shrinks towards zero for large λ. But they never reach zero. They shrunken by quantity 1/(1+λ). Hence,

$$\left|\hat{\beta}_{j_{RR}}\right| \leq \left|\hat{\beta}_j\right| \qquad \text{for j=1,2,… k} \qquad \ldots(3.77)$$

The ridge estimate approaches zero as $\lambda \to \infty$.

The variance of the ridge estimator can be obtained as follows, Let, a linear operator,

$$W = (X'X + \lambda I)^{-1}X'X \quad \ldots(3.78)$$

Then, $$W\hat{\beta} = (X'X + \lambda I)^{-1}X'X(X'X)^{-1}X'y = (X'X + \lambda I)^{-1}X'y = \hat{\beta}_{RR} \quad \ldots(3.79)$$

$$Var(\hat{\beta}_{RR}) = Var(W\hat{\beta}) \quad \ldots(3.80)$$

$$= WVar(\hat{\beta})W'$$

$$Var(\hat{\beta}_{RR}) = \sigma^2 W(X'X)^{-1}W' = \sigma^2(X'X + \lambda I)^{-1}[(X'X + \lambda I)^{-1}X'X\,]' \quad \ldots(3.81)$$

$$Var(\hat{\beta}) - Var(\hat{\beta}_{RR}) = \sigma^2(X'X)^{-1} - \sigma^2 W(X'X)^{-1}W' \quad \ldots(3.82)$$

$$Var(\hat{\beta}) - Var(\hat{\beta}_{RR}) = \sigma^2[(X'X)^{-1} - W(X'X)^{-1}W'] \quad \ldots(3.83)$$

The difference is non negative. The variance of OLS estimator exceeds the variance of ridge estimator.

Also,

$$\lim_{\lambda\to\infty} Var(\hat{\beta}_{RR}) = \lim_{\lambda\to\infty} \sigma^2(X'X + \lambda I)^{-1}[(X'X + \lambda I)^{-1}X'X\,]' = 0 \quad \ldots(3.84)$$

Hence, the variance of the ridge estimator vanishes as ridge penalty approaches infinity.

3.4.2.2 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO regression imposes L_1 penalty in which sum of squared coefficients is replaced by sum of absolute values of coefficient as the penalty term.

$$S(\beta)_{LASSO} = \sum_{i=1}^{n}\left(y_i - \sum x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{k}|\beta_j| \quad \ldots(3.85)$$

An equivalent way to write is

$$S(\beta)_{LASSO} = \|Y - X\beta\|_2^2 + \lambda\|\beta\|_1 \quad \ldots(3.86)$$

Where, the first term in the right-hand side is the RSS function of OLS and $\|\beta\|_1 = \sum_{j=1}^{k}|\beta_j|$.

The idea of LASSO problem is to find a solution to minimize the RSS function subject to the condition that $\sum_{j=1}^{k}|\beta_j| < \eta$. Where, η is any small quantity.

$$S(\beta)_{LASSO} = \|Y - X\beta\|_2^2 + \lambda\|\beta\|_1 \quad \ldots(3.87)$$

$$= \min_{\beta}\left(Y'Y - Y'X\beta - \beta'X'y + \beta'X'X\beta + \lambda\sum|\beta_j|\right)$$

$$= \min_{\beta}\left(-2\beta'\hat{\beta} + \beta'\beta + \lambda\sum|\beta_j|\right) \quad \ldots(3.88)$$

For each coefficient β_j

$$S(\beta)_{LASSO} = \min_{\beta}\left(\sum_{j=1}^{k} -2\beta_j\hat{\beta}_j + \beta_j^2 + \lambda|\beta_j|\right) \quad \ldots(3.89)$$

For each $\beta_j=\beta$ we need to solve the above equation. Hence,

$$S(\beta)_{LASSO} = \min_{\beta}\left(\sum_{j=1}^{k} -2\beta\hat{\beta} + \beta^2 + \lambda|\beta|\right) \quad \ldots(3.90)$$

Case i) If $\hat{\beta} \geq 0$, then the $\hat{\beta}_{LASSO}$ = the value of β that minimizing the $S(\beta)_{LASSO}$ is also greater than 0.

$$\frac{\partial S(\beta)_{LASSO}}{\partial \beta} = -2\hat{\beta} + 2\hat{\beta}_{LASSO} + \lambda \quad \ldots(3.91)$$

Equating this to zero,

$$\frac{\partial S(\beta)_{LASSO}}{\partial \beta} = -2\hat{\beta} + 2\hat{\beta}_{LASSO} + \lambda = 0 \quad \ldots(3.92)$$

$$\hat{\beta}_{LASSO} = \hat{\beta} - \frac{\lambda}{2} \quad \ldots(3.93)$$

Hence if the $\hat{\beta}$ of OLS is positive, then the LASSO estimator is reduced by λ/2.

Case ii) If $\hat{\beta} < 0$, then necessarily $\hat{\beta}_{LASSO} \leq 0$

We have to solve,

$$S(\beta)_{LASSO} = \min_{\beta}\left(\sum_{j=1}^{k} -2\beta\hat{\beta} + \beta^2 - \lambda|\beta|\right) \quad \ldots(3.94)$$

Then,

$$\frac{\partial S(\beta)_{LASSO}}{\partial \beta} = -2\hat{\beta} + 2\hat{\beta}_{LASSO} - \lambda \quad \ldots(3.95)$$

Equating this to zero,

$$\frac{\partial S(\beta)_{LASSO}}{\partial \beta} = -2\hat{\beta} + 2\hat{\beta}_{LASSO} - \lambda = 0 \quad \ldots(3.96)$$

$$\hat{\beta}_{LASSO} = \hat{\beta} + \frac{\lambda}{2} \quad \ldots(3.97)$$

provided that the right-side quantity is negative. Hence, if the $\hat{\beta}$ of OLS is negative, then the LASSO estimator is increased by λ/2.

Thus,

$$\hat{\beta}_{j_{LASSO}} = sign(\hat{\beta}_j)\left(|\hat{\beta}_j| - \frac{\lambda}{2}\right) \quad \ldots(3.98)$$

Further, the LASSO estimates of some of the coefficients are zero provided that their OLS estimates are smaller than 0.5λ.

3.4.2.2 a) Subset selection in LASSO

As a coefficient shrinkage method, ridge regression performance is better due to bias variance trade off. As ridge regression model retains all the predictors in the final model, it is not a parsimonious model.

In best subset selection, for each x_j such that $\{j \in 1,2, \dots k\}$ we select the model of size p that gives the lowest RSS.

i.e. $$\min_{\beta} \sum_{i=1}^{n} \left(Y_i - \sum x_{ij}\beta_j\right)^2 \text{subject to} \sum_{j=1}^{k} 1.(\beta_j \neq 0) \leq p \qquad \dots(3.99)$$

that means we are looking for a set of coefficient estimates that minimizing the SSE subject to have no more than p non zero coefficients among $\{\beta_1, \beta_2, \dots \beta_k\}$.

If the column of independent variable matrix X is orthogonal such that $X'X = I$, then the problem reduced to

$$\min_{\beta} -2\beta'\hat{\beta} + \beta'\beta \qquad \dots(3.100)$$

$$\min_{\beta} \sum_{j=1}^{k} -2\beta_j\hat{\beta}_j + \beta_j^2$$ subject to at most p coefficients.

Hence, we keep only p largest coefficients (in absolute values), that make the sum as small as possible.

i.e. $$\hat{\beta}_{j_{LASSO}} = \hat{\beta}_j(|\hat{\beta}_j| > |\hat{\beta}_{k-p}|) \qquad \dots(3.101)$$

3.4.2.2 b) Shrinkage

Different types of coefficient shrinkage are done in ridge and LASSO regression. In ridge, the constant rate of shrinkage is done for every least square estimate. The LASSO also shrunk the least square estimate at a constant rate unless the absolute value of least square is estimate is less than λ/2. The imposition of quadratic penalty due to ridge had greater impact on the larger coefficients and small penalty for the coefficients near zero. In contrast to that, the sum of absolute penalty due to LASSO, the smaller coefficients skunks quickly towards zero and larger coefficients were less affected by the penalty (Fahrmeir *et al* 2013).

3.4.2.3 Elastic Net (ENET)

Elastic Net was introduced by Zou and Hastie (2005) as another penalized regression model. ENET is combination of Ridge and LASSO. In ENET both sum and absolute penalty are imposed together with RSS function.

$$S(\beta)_{ENET} = \sum_{i=1}^{n} \left(y_i - \sum x_{ij}\beta_j\right)^2 + \lambda \left((1-\alpha) \sum_{j=1}^{p} \left(\beta_j\right)^2 + \alpha \sum_{j=1}^{p} |\beta_j|\right) \qquad \dots(3.102)$$

where, $\lambda > 0$ and $0 < \alpha < 1$. The parameter lambda controls the mix between L_1 and L_2 penalty. It can be seen that the ENET is the generalization of Ridge and Lasso. If $\alpha = 0$, then the RSS function reduces to Ridge and if α=1, then the RSS function reduces to Lasso. if the value of α is near zero the ENET behaves like ridge regression and if α is close to one, then it is of Lasso type. If the α=0.5, then ENET is the equal combination both ridge and Lasso. If λ=0, then the whole penalty vanishes and only the RSS of OLS remains.

An equivalent way to write the RSS is

$$S(\beta)_{ENET} = \|Y - X\beta\|_2^2 + \lambda(1-\alpha)\|\beta\|_2^2 + \lambda\alpha\|\beta\|_1 \quad \dots(3.103)$$

Let, $\lambda(1-\alpha) = \lambda_1$ and $\lambda\alpha = \lambda_2$,

$$\Rightarrow S(\beta)_{ENET} = \|Y - X\beta\|_2^2 + \lambda_1\|\beta\|_2^2 + \lambda_2\|\beta\|_1 \quad \dots(3.104)$$

The idea of ENET problem is similar to LASSO to find a solution to minimize the RSS function. But subject to both the conditions of absolute and sum penalty of coefficients.

$$\min_\beta S(\beta)_{ENET} = \min_\beta(\|Y - X\beta\|_2^2 + \lambda_1\|\beta\|_2^2 + \lambda_2\|\beta\|_1) \quad \dots(3.105)$$

$$= \min_\beta(Y'Y - 2\beta'X'Y + \beta'X'X\beta + \lambda_1\beta'\beta + \lambda_2\sum|\beta|)$$

$$= \min_\beta\left(-2\beta'\hat{\beta} + \beta'\beta + \lambda_1\beta'\beta + \lambda_2\sum|\beta|\right)$$

$$= \min_\beta\left(-2\beta'\hat{\beta} + (1+\lambda_1)\beta'\beta + \lambda_2\sum|\beta|\right) \quad \dots(3.106)$$

Minimization can be done for each variable separately. For each $\beta_j=\beta$ we need to solve the above equation. Hence,

$$= \min_\beta \sum_{j=1}^{k}\left(-2\beta_j\hat{\beta}_j + (1+\lambda_1)\beta_j^2 + \lambda_2\sum|\beta_j|\right) \dots(3.107)$$

Proceed as like Lasso, on $\hat{\beta} \geq 0$, necessarily the solution of $\hat{\beta}_{ENET}$ is positive. Hence the criterion to minimize $-2\beta_j\hat{\beta}_j + (1+\lambda_1)\beta_j^2 + \lambda_2\sum|\beta_j|$ is on right half only. The minimum is obtained as

$$\frac{\partial S(\beta)_{ENET}}{\partial\beta} = -2\hat{\beta} + 2(1+\lambda_1)\hat{\beta}_{ENET} + \lambda_2 \quad \dots(3.108)$$

Equating this to zero,

$$-2\hat{\beta} + 2(1+\lambda_1)\hat{\beta}_{ENET} + \lambda_2 = 0$$

$$-2\hat{\beta} + 2(1+\lambda_1)\hat{\beta}_{ENET} = -\lambda_2$$

$$2\left[(1+\lambda_1)\hat{\beta}_{ENET} - \hat{\beta}\right] = -\lambda_2$$

$$(1+\lambda_1)\hat{\beta}_{ENET} - \hat{\beta} = \frac{-\lambda_2}{2}$$

$$(1+\lambda_1)\hat{\beta}_{ENET} = \hat{\beta} - \frac{\lambda_2}{2}$$

$$\hat{\beta}_{ENET} = \frac{1}{(1+\lambda_1)}\left(\hat{\beta} - \frac{\lambda_2}{2}\right) \quad \ldots(3.109)$$

On substituting $\lambda_1 = \lambda(1-\alpha)$ and $\lambda_2 = \lambda\alpha$

$$\hat{\beta}_{ENET} = \frac{1}{[1+\lambda(1-\alpha)]}\left(\hat{\beta} - \frac{\lambda\alpha}{2}\right) \quad \ldots(3.110)$$

Hence, minimum is obtained at this point provided this quantity is positive and it is at zero otherwise.

Similarly, if $\hat{\beta} < 0$, the solution of $\hat{\beta}_{ENET}$ is negative. The criterion to minimize $-2\beta_j\hat{\beta}_j - (1+\lambda_1)\beta_j^2 - \lambda_2 \sum\left|\beta_j\right|$ is on left half. The minimum is obtained as

$$\hat{\beta}_{ENET} = \frac{1}{(1+\lambda_1)}\left(\hat{\beta} + \frac{\lambda_2}{2}\right) \quad \ldots(3.111)$$

Thus,

$$\hat{\beta}_{ENET} = \frac{1}{(1+\lambda_1)} sign(\hat{\beta})\left(\left|\hat{\beta}\right| - \frac{\lambda_2}{2}\right) \quad \ldots(3.112)$$

or

$$\hat{\beta}_{ENET} = \frac{1}{[1+\lambda(1-\alpha)]} sign(\hat{\beta})\left(\left|\hat{\beta}\right| - \frac{\lambda\alpha}{2}\right) \quad \ldots(3.113)$$

3.4.3 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is the most popular and widely used machine learning technique to model the real-world problem. The ANN became a promising alternative to traditional statistical models for both classification as well as prediction context. As the ANN models were non-parametric in nature, there was no prior assumption to be made on the distribution of data.

The concept of ANN was inspired from the learning mechanism of biological organisms. The basic signalling units of human nervous system are referred as neurons which are connected to each other through dendrites and axons. The connection between one neuron to another neuron is termed as synapse. The signal or information flow from one neuron to another neuron is done through the use of these interconnections. The ANN algorithm mimics this process.

The building blocks of ANN is a neuron or node which is a simple computation unit. A typical ANN algorithm is organized into layers of these neuron. Each layer is a row of one or more neurons. The connection is made in such a way to connect the nodes of successive layers. The amount of input received to node a layer from a node from previous layer is depends on the strength of connection between the nodes which is termed as weight. These wights were adjustable to strengthen or weaken an input.

3.4.3.1 Multilayer Perceptron (MLP)

The processing units of the ANN is connected each other according to certain networking topology. Multilayer Perceptron (MLP) is the most commonly used type of Feedforward Neural Network topology. Perceptron is a single neuron or node model which is a precursor to high complex neural network. The MLP architecture is a fully connected neural net architecture where each neuron of a layer is connected with every neuron of previous and/or next layer.

A typical MLP architecture consists of three layers namely, input layer, hidden layer and an output layer. The input layer obtains input data (x_i) into the neural network system and an appropriate weight (w_i) is to be multiplied with each input. A bias term (b) also incorporated as weight of an edge using a bias neuron. The sum of this product is termed as activation value which is transmitted to the hidden layer. In hidden layer a nonlinear transformation is to be applied through a nonlinear activation function. This computed value transmitted to the output layer as a linear combination of all the neurons of hidden layer. This structure learns nonlinear relationships between input and output vectors.

The relationship between output and inputs can be mathematically represented as follows:

$$Y_t = f\left\{\sum_{j=1}^{q} \omega_j . g\left(\sum_{i=1}^{k} (w_i . x_i) + b . w_0\right)\right\} \qquad \text{... (3.114)}$$

where, k is the number of input nodes, q is the number of hidden nodes, ω_i (i=1,2,…k) are connection weights of input layer, ω_j (j=1,2,…q) are output layer weights, b is bias term and g and f denote the activation function at hidden and output layer respectively and e_t is error term. The graphical representation of ANN with MLP topology is given in the Fig.3.2.

The MLP algorithm approximate output as nonlinear function of input variables even when there is no assumption has made on the nature relationship between input and output variables. This is the advantage of MLP over the linear regression model to capture the nonlinearity while fitting a model for predicting a response variable using many input variables.

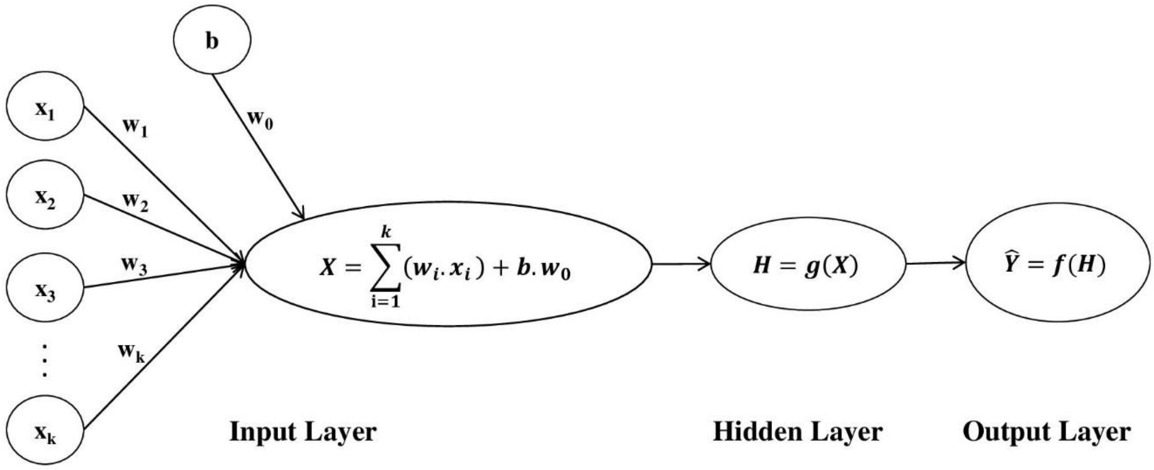

Fig. 3.2 Graphical representation of MLP topology of ANN

3.4.3.2 Back-propagation (BP) algorithm

The process of finding the right value of weight for each input is called as learning rule. The procedure to adjust the weight in learning rules is termed as learning algorithm (Noriega, 2005). Many training algorithms are being used to train the ANN model. As the feedforward MLP are memoryless in the sense its output for a given input during a specified training run is independent of its previous trainings. The main drawback of feedforward MLP is the information flow is in unidirectional. Hence, the weights are fixed.

Hence, Rumelhart (1986) found an algorithm to update the weights in feedforward MLP which was termed as Backpropagation (BP) learning law or Generalized Delta rule. The BP algorithm starts training the input data with random weights and the weights are adjusted in successive steps to reduce error. BP algorithm uses the error signal for determining, to what degree of weights to be adjusted to reduce overall error of MLP training. The BP algorithm uses gradient descent algorithm on the error surface to obtain optimum set of weights.

Let, $I = \sum_{i=1}^{k}\left(w_{im}^{h}.x_i\right)$ is the sum of product of an input x_i and its corresponding initial weight w_{im}^{h} plus bias term to the m^{th} hidden layer. Let, w_{m}^{o} is the initial weight between m^{th} hidden layer and the output layer neuron. Both the initial weights were to be adjusted by learning of specific input-output patterns over many steps iteratively in order to reduce the gap between actual and predicted output. Let, I(t) and Y(t) were the input-output pattern at t^{th} steps. $\hat{Y}(t)$ is the output that is obtained through the output layer for input-output pattern I(t) and Y(t) at the t^{th} steps. Y(t) is the actual output value.

The Mean Square Error at the t^{th} steps is given as

$$MSE(t) = E(t) = \frac{1}{2}\sum_{j=1}^{n}\left[Y(t) - \hat{Y}(t)\right]^2 \qquad \ldots(3.115)$$

$$= \frac{1}{2}\sum_{j=1}^{n}[Y(t) - f(w_{ml}^{o}(t).H_{ml})]^2 \quad \ldots(3.116)$$

where,

$$H_m = g(I_{im}) \quad \ldots(3.117)$$

$$I_{im} = \sum_{\mathrm{i=1}}^{\mathrm{k}}\left[w_{im}^{h}(t).\mathrm{x_{im}}\right] \quad \ldots(3.118)$$

The weight w_m^o between m[th] hidden layer neuron and the output neuron in t+1 step is

$$w_{ml}^{o}(t+1) = w_{ml}^{o}(t) + \Delta w_{ml}^{o}(t) \quad \ldots(3.119)$$

where,

$$\Delta w_{ml}^{o}(t) = -\eta\frac{\partial E(t)}{\partial w_{ml}^{o}(t)} \quad \ldots(3.120)$$

η is the learning rate.

Using chain rule, $$O_l = \sum \Delta w_{ml}^{o}(t).H_{ml} \quad \ldots(3.121)$$

$$\hat{Y}_l = f(O_l) \quad \ldots(3.122)$$

$$\frac{\partial E(t)}{\partial w_{ml}^{o}(t)} = \frac{\partial E(t)}{\partial O_l}\frac{\partial O_l}{\partial w_{ml}^{o}(t)}$$

As, $$O_l = \sum \Delta w_{ml}^{o}(t).H_{ml} \quad \ldots(3.123)$$

$$\frac{\partial O_l}{\partial w_{ml}^{o}(t)} = H_{ml} \quad \ldots(3.124)$$

Hence,

$$\frac{\partial E(t)}{\partial w_{ml}^{o}(t)} = \frac{\partial E(t)}{\partial O_l}H_{ml} \quad \ldots(3.125)$$

Here,

$$\frac{\partial E(t)}{\partial O_l} = \frac{\partial E(t)}{\partial \hat{Y}_l}.\frac{\partial \hat{Y}_l}{\partial O_l} \quad \ldots(3.126)$$

$$\frac{\partial E(t)}{\partial \hat{Y}(t)} = \frac{\partial}{\partial \hat{Y}(t)}\left\{\frac{1}{2}\sum_{j=1}^{n}\left[Y - \hat{Y}_l\right]^2\right\} = -\left[Y - \hat{Y}_l\right] \quad \ldots(3.127)$$

$$\frac{\partial \hat{Y}(t)}{\partial O_l} = \frac{\partial[f(O_l)]}{\partial O_l} = \hat{Y}_l\left[1 - \hat{Y}_l\right] \quad \ldots(3.128)$$

On substitution of (3.128) into (3.126),

$$\frac{\partial E(t)}{\partial O_l} = -\hat{Y}_l[Y - \hat{Y}_l][1 - \hat{Y}_l] \qquad \text{... (3.129)}$$

On substitution of (3.129) into (3.125),

$$\frac{\partial E(t)}{\partial w_{ml}^o(t)} = -\hat{Y}_l[Y - \hat{Y}_l][1 - \hat{Y}_l]H_{ml} \qquad \text{... (3.130)}$$

Finally on substitution of (3.130) into (3.120),

So, $$\Delta w_m^o(t) = \eta\hat{Y}_l[Y - \hat{Y}_l][1 - \hat{Y}_l]H_{ml} \qquad \text{...(3.131)}$$

Let, $$\delta_l = \hat{Y}_l[Y - \hat{Y}_l][1 - \hat{Y}_l] \qquad \text{...(3.132)}$$

Then, $$\Delta w_{ml}^o(t) = \eta.\delta_l H_{ml} \qquad \text{...(3.133)}$$

Hence, the new weight between mth hidden neuron and the lth output neuron in t+1 step $w_{ml}^o(t+1)$ is

$$w_{ml}^o(t+1) = w_m^o(t) + \eta.\delta_l H_{ml} \qquad \text{...(3.134)}$$

Similarly, the weight w_m^o between mth hidden layer and output neuron in t+1 step is

$$w_{im}^h(t+1) = w_{im}^h(t) + \Delta w_{im}^h(t) \qquad \text{...(3.135)}$$

$$\Delta w_{im}^h(t) = -\eta\frac{\partial E(t)}{\partial w_m^o(t)} \qquad \text{...(3.136)}$$

$$\frac{\partial E(t)}{\partial w_m^o(t)} = \frac{\partial E(t)}{\partial H_{im}}\frac{\partial H_{im}}{\partial w_m^o(t)} \qquad \text{.... (3.137)}$$

As, $$H_{im} = \sum[w_{im}^h(t).\mathrm{x_{im}}] \qquad \text{...(3.138)}$$

$$\frac{\partial H_{im}}{\partial w_m^o(t)} = \mathrm{x_{im}} \qquad \text{... (3.139)}$$

Hence,

$$\frac{\partial E(t)}{\partial w_m^o(t)} = \frac{\partial E(t)}{\partial H_{im}}\mathrm{x_{im}} \qquad \text{... (3.140)}$$

Using chain rule,

$$\frac{\partial E(t)}{\partial H_{im}} = \sum\frac{\partial E(t)}{\partial O_l}\frac{\partial O_l}{\partial H_{im}} \qquad \text{... (3.141)}$$

Let,

$$\frac{\partial E(t)}{\partial H_{im}} = \sum -\delta_l \frac{\partial O_l}{\partial H_{im}} \qquad \text{... (3.142)}$$

$$= -\sum \delta_l \frac{\partial O_l}{\partial \hat{Y}_l} \frac{\partial \hat{Y}_l}{\partial H_{im}}$$

$$= -\sum \delta_l\, w_{ml}^o \frac{\partial \hat{Y}_l}{\partial H_{im}}$$

$$\frac{\partial E(t)}{\partial H_{im}} = -\sum \delta_l\, w_{ml}^o . \hat{Y}_l\left[1 - \hat{Y}_l\right] \qquad \text{... (3.143)}$$

On substitution of (3.143) and (3.139) into (3.137),

$$\frac{\partial E(t)}{\partial w_m^o(t)} = -\hat{Y}_l\left[1 - \hat{Y}_l\right] \sum \delta_l\, w_{ml}^o . \mathrm{x_{im}} \qquad \text{... (3.144)}$$

On substitution of (3.144) into (3.136),

$$\Delta w_{im}^h(t) = \eta \hat{Y}_l\left[1 - \hat{Y}_l\right] \sum \delta_l\, w_{ml}^o . \mathrm{x_{im}} \qquad \text{... (3.145)}$$

Let, $$\delta_m = \hat{Y}_l\left[1 - \hat{Y}_l\right] \sum \delta_l\, w_{ml}^o \qquad \text{...(3.146)}$$

Then, $$\Delta w_{im}^h(t) = \eta \delta_m \mathrm{x_{im}} \qquad \text{...(3.147)}$$

Hence, the new weight between i[th] input neuron and the m[th] hidden neuron in t+1 step $w_{im}^h(t+1)$ is

$$w_{im}^h(t+1) = w_{im}^h(t) + \eta \delta_m \mathrm{x_{im}} \qquad \text{...(3.148)}$$

The BP algorithm avoids to converge to a local minimum with high speed of convergence while training the ANN model (Iranmanesh and Mahdavi, 2009). The learning rate η decides the weight adjustment by scaling gradient of error surface. For a small learning rate, the adjustment in the weight is also small. As a result of which the learning process is slow that takes many iterations to converge. Too large learning rate speeds up the learning process by making huge changes in the weights in each single step that leads to the oscillatory and unstable trajectory on error surface. A compromise between these extremes can be obtained by implementing backpropagation using adaptive learning rate (Chaturvedi, 2008). In our study, the learning rate ranges between 0.01 and 1.00 have been tried at 0.01 interval. Hence 100 different learning rates are under consideration. The optimum learning rate was chosen for the value of η at which the error function of gradient descent algorithm converged to the minimum.

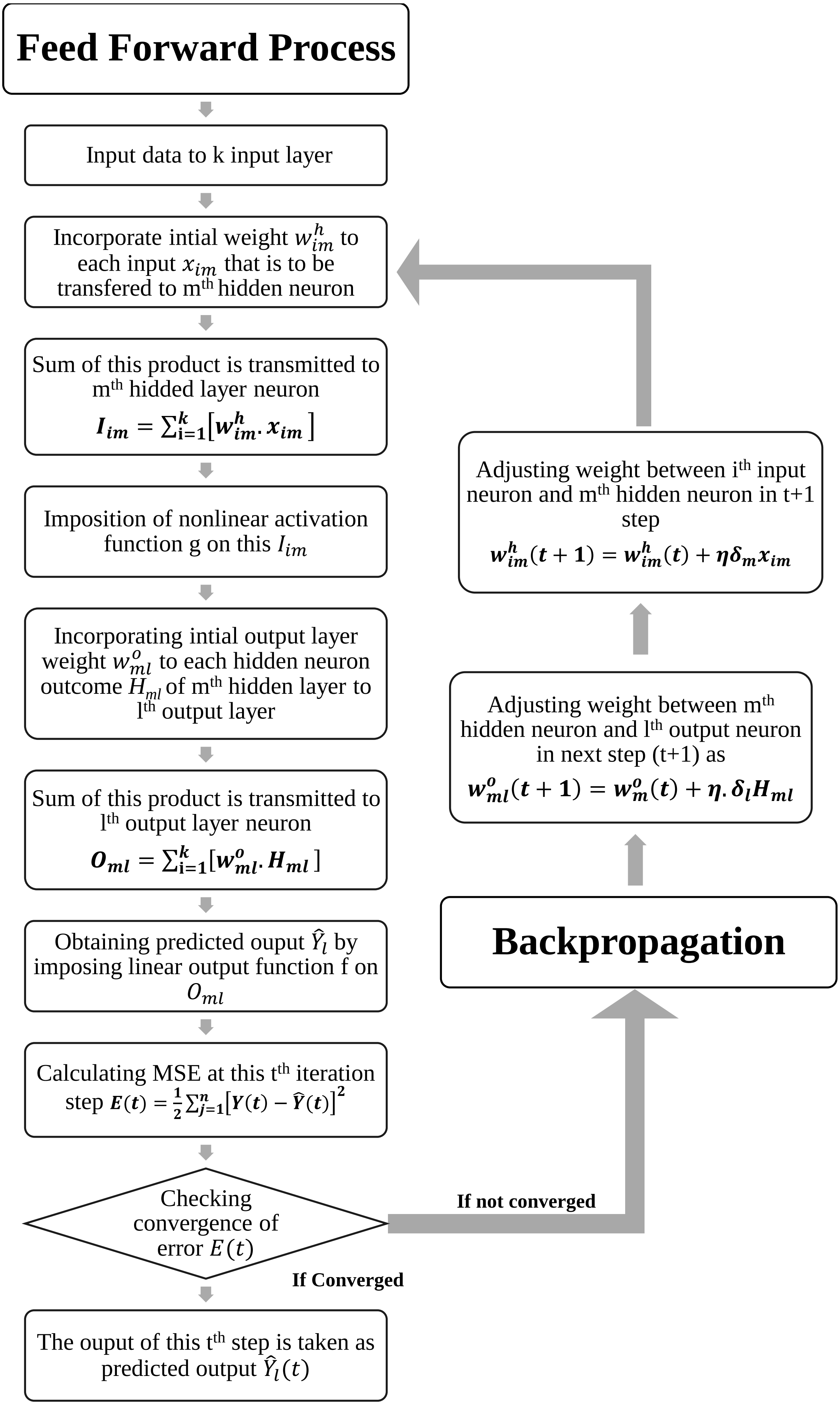

Fig. 3.3 Two-step training procedure of Multilayer Perceptron (MLP) using Backpropagation (BP)

3.4.3.3 Optimizing number of hidden layer neurons

The choice of activation function and number of neurons in the hidden layer are the critical part in the neural network design (Aggarwal, 2018). A MLP that has a single hidden layer with appropriate number of neurons coupled with right choice of nonlinear activation function is able to capture nonlinear association between input and output variables.

The BP algorithm converges to a local minimum rather than global minimum or may not be converged to minimum when there is a less number of neurons in the hidden layer. It may cause underfitting of model and more statistical bias as well. In converse to this, too many hidden layer neurons generally lead to overfitting, high variation and more computational time due to over complexed model (Ke and Liu, 2008). Hence, determination of optimal number of hidden layer neuron is necessary for fitting optimally performing neural network model (Sheela and Deepa, 2013).

The number of the hidden layer neurons is optimised by 10-fold cross validation procedure (Srinivasan *et al* 2019). The number of hidden layer neurons from 1 to 10 are evaluated. The optimum number of neurons in hidden layer is the number of neurons for which the Root Mean Square Error for Cross validation (RMSE-CV) is low.

3.4.3.4 Activation Function

Activation Function or Transfer Function is a mathematical function that is responsible for the nonlinearity in neutral network. The linear activation function in the hidden layer would be a linear approximation of input values. The nonlinear activation function maps the weighted sum of input data to a nonlinear output. An appropriate nonlinear activation function learns the pattern of nonlinearity that exists in the training dataset and it improves the learning rate as well. Without a nonlinear activation function, the hidden layer neuron performs a linear transformation on weighted sum of input and bias. The increased computational power of the MLP can be achieved by the proper choice of non-linear activation function in the hidden layer.

The right choice of activation function is context dependent since no activation function performs exceptionally best in every context (Sharma *et al* 2018). In this present study, the performance of MLP with backpropagation algorithm using three nonlinear hidden layer activation functions *viz.,* Logistic, Tangent hyperbolic (Tanh) and Restricted Linear Unit (ReLU) have been evaluated.

3.4.3.4 a) Logistic Function

The logistic function is the most commonly used activation function on neural network. Mathematically, the logistic activation function is given as,

$$g(x) = \frac{1}{(1+e^{-x})} \qquad \ldots(3.149)$$

where $x \in (-\infty, +\infty)$ and $g(x) \in (0,1)$. It is a differentiable and monotonic function. Since the function is not symmetrical about zero, the signs of all output from logistic is same. The value of g(x) is close to one for larger input and it is close to zero for smaller input. This relationship is represented by a Sigmoid (S) shaped curve.

3.4.3.4 b) Tangent hyperbolic (Tanh) Function

The Tanh activation function is also S shape curve function similar to logistic function but its output ranges between -1 and +1. The functional relation of Tanh function is given as,

$$g(x) = \frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})} \qquad \ldots(3.150)$$

Unlike the sigmoid function, the Tanh function is zero centred since the output can be mapped into positive as well as negative. The larger the input, the output will be near +1 and the output moves towards -1 for smaller input. As the output ranges between -1 to +1, the average output from hidden layer will be zero or near zero. This cantering of output from hidden layer makes the output layer learn much easier. As the gradient of Tanh function is much steeper than sigmoid function, the former is always better than later in terms of nonlinearity.

3.4.3.4 c) Restricted Linear Unit (ReLU) Function

The ReLU functions are linear for positive inputs and zero for negative inputs.

$$g(x) = max(0, x) = \begin{cases} x_i \; if \; x_i \geq 0 \\ 0 \; if \; x_i < 0 \end{cases} \qquad \ldots(3.151)$$

The ReLU function does not activate all neurons at same time. As only certain neurons were activated, the computational efficiency of ReLU is more than sigmoid and Tanh functions. The weights of some neurons are not updated during the backpropagation since the neuron output becomes zero for the negative input values (Madhu *et al* 2023). This leads to some dead neuron. As a result of which the model fit and generalized performance of the model get affected. But linearity for the positive input has an attractive property of non-saturation of gradient even though some of gradient are zero.

3.4.3.4 d) Linear Activation Function for output layer

Linear activation function which is also known as identity function is a function in which the activation is proportional to it input. The linear function does not do anything on the weighted sum. It just passes the input as like linear regression. Mathematically, the linear activation function is,

$$f(x) = x \quad \ldots(3.152)$$

While fitting an Artificial Neural Network (ANN) model to any real-world data, the questions to be answered are which Network architecture to be used, the Learning algorithm to update neural weights, the methodology to fix the number of hidden layer neurons and activation function to be used in hidden layer (Hunter *et al* 2012). In our present study, Multilayer Perceptron (MLP) architecture of ANN is used with Backpropagation (BP) to update neural weights. 10-fold cross validation procedure is employed to determine the number of hidden layer neurons. The performance of three activation functions viz. Logistic, Tangent hyperbolic (Tanh) and Restricted Linear Unit (ReLU) are evaluated.

Data normalization is necessary before training ANN model to avoid bias due to magnitude of input variable and its range of variation (Puig-Arnavat and Bruno, 2015). Many numbers of input variables resulting more complex network of neuron as well as many weights to be adjusted. In consequence of these, the convergence speed of the BP algorithm gets slow downed (Gardner and Dorling, 1998) Removing of redundant variables and dimension reduction are the solutions to avoid the problem of slow convergence of network algorithm.

3.4.4 Support Vector Regression (SVR)

Support vector machines (SVM) is a robust and efficient nonlinear multivariate machine learning tool. Initially, the SVM was developed by Vapnik for classification problems by formulating convex optimizing problem. The SVM maps the input feature vectors into a high dimensional space that creates a best separating hyperplane (Nti *et al* 2021). This hyperplane is represented by support vectors.

For a continuous valued output, the SVM can be generalization to the regression context which was termed as Support Vector Regression (SVR). The basic concept of SVR is to transform the input data into a high dimensional feature space and a linear regression is to be constructed in the new high dimensional space which corresponds to a nonlinear regression in the original input space.

The optimization problem in SVR is formulated by introducing a ε-insensitive tube or hyperplane. The hyperplane that best approximates the continuous valued output by balancing both model complexity and prediction error has to identified. The optimization procedure in SVR is formulated by defining a convex ε-insensitive loss function.

Let y is the dependent variable and x is a set of input data, the support vector regression is expressed as,

$$y = f(x) = \sum_{i=1}^{k} w_i x_i + b \quad \ldots(3.153)$$

This can be written in matrix notation as follows,

$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \mathcal{W}^T X + b \qquad \ldots(3.154)$$

The optimization problem in SVR attempts to find the narrowest hyperplane centred around the surface and minimize the difference between actual and predicted output as well. The optimization problem produces the following objective function,

$$Min_w \frac{1}{2} \|\mathcal{W}\|^2 \qquad \ldots(3.155)$$

The geometric properties of the hyperplane are the major concern in SVR model. The data points that lie on and outside around the hyperplane are called as support vectors (Awad and Khanna, 2015). The support vectors and width of hyperplane (ε) are the most important factors that determine the shape of hyperplane. Smaller value of ε, lowers the tolerance level for error as well as number of support vectors. The margin of the hyperplane moves inwards if the width of hyperplane (ε) is decreased. This leads to the occurrence of more datapoint around the margin or more number of support vector. Increasing ε results into fewer support vectors. The aim of SVR to contain all the datapoints within margin of the hyperplane (ε). The loss function is to be minimized in order to find a flattest hyperplane that contains most of the training data points.

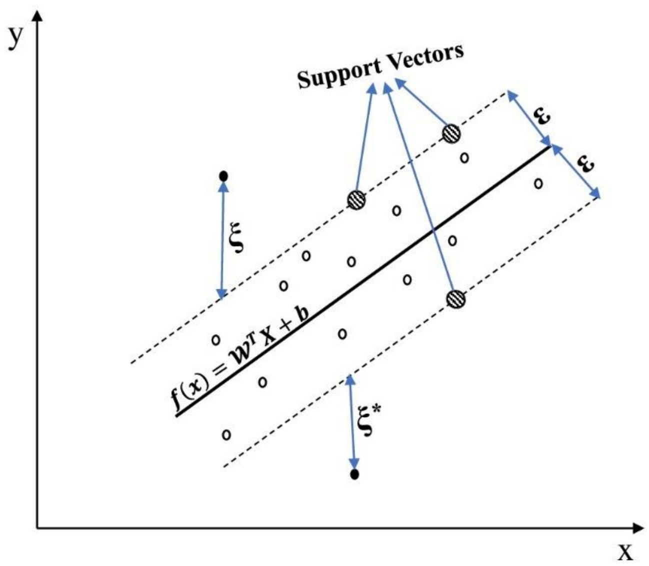

Fig. 3.4 Graphical representation of SVR

The ε-insensitive tube makes the SVR more robust as it is less sensitive for noisy inputs. In order to guard against outliers, the soft margin approach is to applied by including slack variables ξ and ξ* in the optimizing function (Cortes and Vapnik, 1995) as follows,

$$Min \frac{1}{2} \|\mathcal{W}\|^2 + C \sum_{i=1}^{N} \xi_i + \xi_i^* \qquad \ldots(3.156)$$

Subject to

$$\left.\begin{aligned} y_i - \mathcal{W}^T x_i \leq \varepsilon + \xi_i^* \\ \mathcal{W}^T x_i - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{aligned}\right\} \forall \text{ i} = 1,2, \dots, \text{N} \qquad \dots(3.157)$$

The ξ and ξ* represent the distance from the datapoints that lies outside the tube to the upper and lower margins of the hyperplane respectively. The ξ and ξ* equals to zero for the datapoints that lie inside the hyperplane as well as on the margin of the hyperplane (support vectors). The slack variables determine the number of points lie outside the tube can be tolerated. The ε-loss function penalizes the datapoints that are farther than the margin of the hyperplane (ε) from the centre.

C is the regularization parameter or cost. The value C>0, determines the trade-off between size of the hyperplane and the maximum distance of datapoints that are larger than ε can be tolerated. C is a tuneable parameter which gives weightage to minimize the tube flatness. A larger C gives more weightage to the minimize error but its generalized performance is generally poor for new data during validation stage. A very low C imposes less penalty for error

The constraint optimization problem can be solved by writing the equation in Lagrangian form as given below,

$$\mathcal{L}(\mathcal{W}, \xi, \xi^*, \lambda, \lambda^*, \alpha, \alpha^*) = \frac{1}{2}\|\mathcal{W}\|^2 + C\sum_{i=1}^{N} \xi_i + \xi_i^* + \sum_{i=1}^{N} \alpha_i^*(y_i - \mathcal{W}^T x_i - \varepsilon - \xi_i^*) + \sum_{i=1}^{N} \alpha_i(-y_i + \mathcal{W}^T x_i - \varepsilon - \xi_i) - \sum_{i=1}^{N} \lambda_i \xi_i + \lambda^*{}_i \xi_i^* \qquad \dots(3.158)$$

where, λ, λ*, α and α* are the Lagrange multipliers which are non-negative real numbers.

The minimum of the Lagrangian form equation can be solved using Karush-Kuhn-Tucker (KKT) conditions by taking partial derivatives with respect to each variable and equating them to zero as follows,

$$\left.\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathcal{W}} = \mathcal{W} - \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)\, x_i = 0 \\ \frac{\partial \mathcal{L}}{\partial \xi_i^*} = c - \lambda^*{}_i - \alpha_i^* = 0 \\ \frac{\partial \mathcal{L}}{\partial \xi_i} = c - \lambda_i - \alpha_i = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda^*{}_i} = \sum_{i=1}^{N} \xi_i^* \leq 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_i} = \sum_{i=1}^{N} \xi_i \leq 0 \\ \frac{\partial \mathcal{L}}{\partial \alpha_i^*} = y_i - \mathcal{W}^T x_i - \varepsilon - \xi_i^* \leq 0 \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} = -y_i + \mathcal{W}^T x_i - \varepsilon - \xi_i \leq 0 \end{aligned}\right\} \qquad \dots(3.159)$$

These above equations result into the following equations,

$$\left.\begin{aligned} \alpha_i^*(y_i - \mathcal{W}^T x_i - \varepsilon - \xi_i^*) = 0 \\ \alpha_i(-y_i + \mathcal{W}^T x_i - \varepsilon - \xi_i) = 0 \\ \lambda^*{}_i \xi_i^* = 0 \\ \lambda_i \xi_i = 0 \end{aligned}\right\} \forall i \qquad \ldots(3.160)$$

On solving the above equations,

$$\mathcal{W} = \sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i)\, x_i \qquad \ldots(3.161)$$

On substitution,

$$f(x) = \sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i) x_i^T\, x_i \quad \alpha_i^*, \alpha_i \in [0, C] \qquad \ldots(3.162)$$

The datapoints that lie inside the hyperplane are corresponded by the Lagrange multipliers which are equal to zero. The support vectors or datapoints that exist outside are represented by non-zero Lagrange multipliers. The datapoints that satisfies $(\alpha_i^* - \alpha_i) \neq 0$ are the support vectors.

The above expression of f(x) corresponds to a linear function. In order to solve a nonlinear relationship with high accuracy, the input data can be transformed to a high dimensional space using a kernel function. The optimizing function can be expressed by replacing x_i with $\varphi(x_i)$,

$$Min \frac{1}{2} \|\mathcal{W}\|^2 + C \sum_{i=1}^{N} \xi_i + \xi_i^* \qquad \ldots(3.163)$$

Subject to

$$\left.\begin{aligned} y_i - \mathcal{W}^T \varphi(x_i) \leq \varepsilon + \xi_i^* \\ \mathcal{W}^T \varphi(x_i) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{aligned}\right\} \forall \text{ i} = 1,2, \ldots, \text{N} \qquad \ldots(3.164)$$

where, $\varphi(x_i)$ is the transformation from feature space to kernel space. The Lagrangian form for the transformed form is written as

$$\mathcal{L}(\mathcal{W}, \xi, \xi^*, \lambda, \lambda^*, \alpha, \alpha^*) = \frac{1}{2} \|\mathcal{W}\|^2 + C \sum_{i=1}^{N} \xi_i + \xi_i^* + \sum_{i=1}^{N} \alpha_i^* (y_i - \mathcal{W}^T \varphi(x_i) - \varepsilon - \xi_i^*) + \sum_{i=1}^{N} \alpha_i (-y_i + \mathcal{W}^T \varphi(x_i) - \varepsilon - \xi_i) - \sum_{i=1}^{N} \lambda_i \xi_i + \lambda^*{}_i \xi_i^*$$

…(3.165)

The new weight vector in terms of transformed input is,

$$\mathcal{W} = \sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i)\, \varphi(x_i) \qquad \ldots(3.166)$$

where N_{SV} represents the number of support vectors. Finally, the output function in terms of kernel function is

$$f(x) = \sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i)\, K(x_i, x_k) \qquad \ldots(3.167)$$

where, $K(x_i, x_k)$is the kernel function that maps two-vector input space vector x_i and x_k into high dimensional feature space such that $K(x_i, x_k) = \varphi(x_i).\varphi(x_k)$. From the above equation, it is evident that the output is independent of input dimension and is depends on the new high dimensional feature space and the number of support vectors (Smola and Schölkopf, 2004).

3.4.4.1 Kernels

The kernel is a function which simulates the input data into higher dimension feature space (Yekkehkhany *et al* 2014). The performance of SVR to a nonlinear problem is strongly determined by the proper choice of kernel function. Any symmetric function in original input space that is capable of representing a scalar product in higher dimensional feature space can be used as the kernel function, *i.e.* $K(\mathrm{x_i}, \mathrm{x_k}) = \varphi(\mathrm{x_i}).\varphi(\mathrm{x_k})$.

Linear kernel function can be used when the data is linearly separated. Nonlinear kernel such as Radial Basis Function (RBF) and polynomial kernel were appropriate when the data is not linearly separated (Al-Azies *et al* 2019). The performance SVR model under three kernel functions are evaluated in this study. They are,

Linear kernel**:** $k(x_i, x_k) = x_i^T.x_k$ …(3.168)

Polynomial kernel: $k(x_i, x_k) = (\gamma.x_i^T.x_k + 1)^d$ …(3.169)

Radial Basis Function (RBF) kernel: $k(x_i, x_k) = exp(-\gamma\|x_i - x_k\|^2)$ …(3.170)

3.4.4.2 Hyperparameters Tuning

Hyperparameter in SVR are the values that control the learning process. The hyperparameters are to be tuned to the optimum level that minimize the ϵ-loss function. The SVR parameters C and ε along with kernel parameters have to optimized for the generalized performance of the model.

The Cost (C) and Epsilon (ε) are the hyperparameters to be optimized while training SVR model using linear kernel. Along with Cost and Epsilon, an additional parameter Gamma (γ) has to be optimized in nonlinear RBF kernel function. Another parameter, degree of polynomial (d) has to be optimized in polynomial kernel function. Cost (C) adds penalty term to the optimization problem that acts as a leverage between width of hyperplane margin and training error, Epsilon (ε) is the margin of the hyperplane and Gamma (γ) is the scale parameter (Savas and Dovis, 2019). The degree of polynomial is of one is similar to linear kernel. Higher degree of polynomial is required for explaining non-linear relationship between features (Hasan *et al* 2016).

If the C increases infinitely large, the SVR model does not allow error. The error would be more if the C reduces to zero. If ε is very large, there would be few support vectors and a smaller ε results more support vectors that leads to the chance of overfitting (Samui *et al* 2008). Fixing the value of ε at quite low level and setting C and γ as high as is sensible which gives probably the best fit (Brereton and Lloyd, 2010).

In the present study, grid search has been done using 10-fold cross validation to find the right combination of hyperparameters (Pence *et al* 2022). The range of cost (C) parameter is set as 1 to 32, both epsilon (ϵ) and gamma (ϒ) are set from 0 to 1 with an interval of 0.1. The degree of polynomial (d) is the fourth parameter in polynomial kernel which was tuned from 1 to 10.

3.4.5 Model Evaluation Criteria

A traditional Multiple Linear Regression (MLR), three penalized regression models *viz.* Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) and two machine learning model such as Artificial Neural Network (ANN) and Support Vector Regression (SVR) have been fitted for each crop in each district using all the fifteen weather indices as independent or input variables and respective crop yield as dependent variable.

Further MLR, ANN and SVR models have been fitted under three variable selection or dimension reduction techniques namely Stepwise Regression (SR), Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR) in a view of avoiding overfitting issue, increasing speed of training algorithm, to avoid multicollinearity issue as well as to get a less complex model that enables easy interpretation. As a result of which the performance of each of these models can also be examined under three variable selection methodologies.

Since the penalized regression models are doing inner variable selection by shrinking the coefficient values as well as these models mitigate the problem of multicollinearity by their parameter estimation procedure, there is no necessity of variable selection or dimension reduction.

Hence twelve models have been fitted for each crop in each district. Among the fitted models, the best fitted model for each crop in each district has been selected by examining following model evaluation criteria.

Coefficient of Determination: $$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} \quad \ldots(3.171)$$

Mean Absolute Error: $$MAE = \frac{\sum_{i=1}^{n}\left|\left(Y_i - \hat{Y}_i\right)\right|}{n} \quad \ldots(3.172)$$

Root Mean Square Error: $$RMSE = \sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \big/ n} \quad \ldots(3.173)$$

normalized Root Mean Square Error: $\text{nRMSE} = \frac{\text{RMSE}}{\bar{Y}}$ …(3.174)

where, Y_i and $\hat{Y}_i$ are observed and predicted yield corresponding to i^{th} year and $\bar{Y}_i$ is the average yield.

3.4.6 Model Validation and Forecasting

The data from the year 1997-98 to 2019-20 have been utilized to fit the models and the data of the year 2020-21 is reserved to compare the actual and forecasted yield. Validation of the model is essential to check the generalized performance of the model for the new data which are not utilized while training or fitting the model. Hence 80% of the data between the years 1997-98 and 2019-20 are randomly selected for model training. Remaining 20% of data are utilized for validation of the fitted models using residual measures such as MAE, RMSE and nRMSE. The random selection is employed in order to ensure the presence of recent as well as past year's data in both training and testing data set.

The model which performs better in both training and validation stages are considered as the best fitted model. The best fitted model for each crop in each district is selected. The yield is forecasted for the year 2020-21. The forecasted yield is compared with actual yield. Further, the yield of all the five crops in each of the four districts are forecasted for the years 2021-22 and 2022-23.

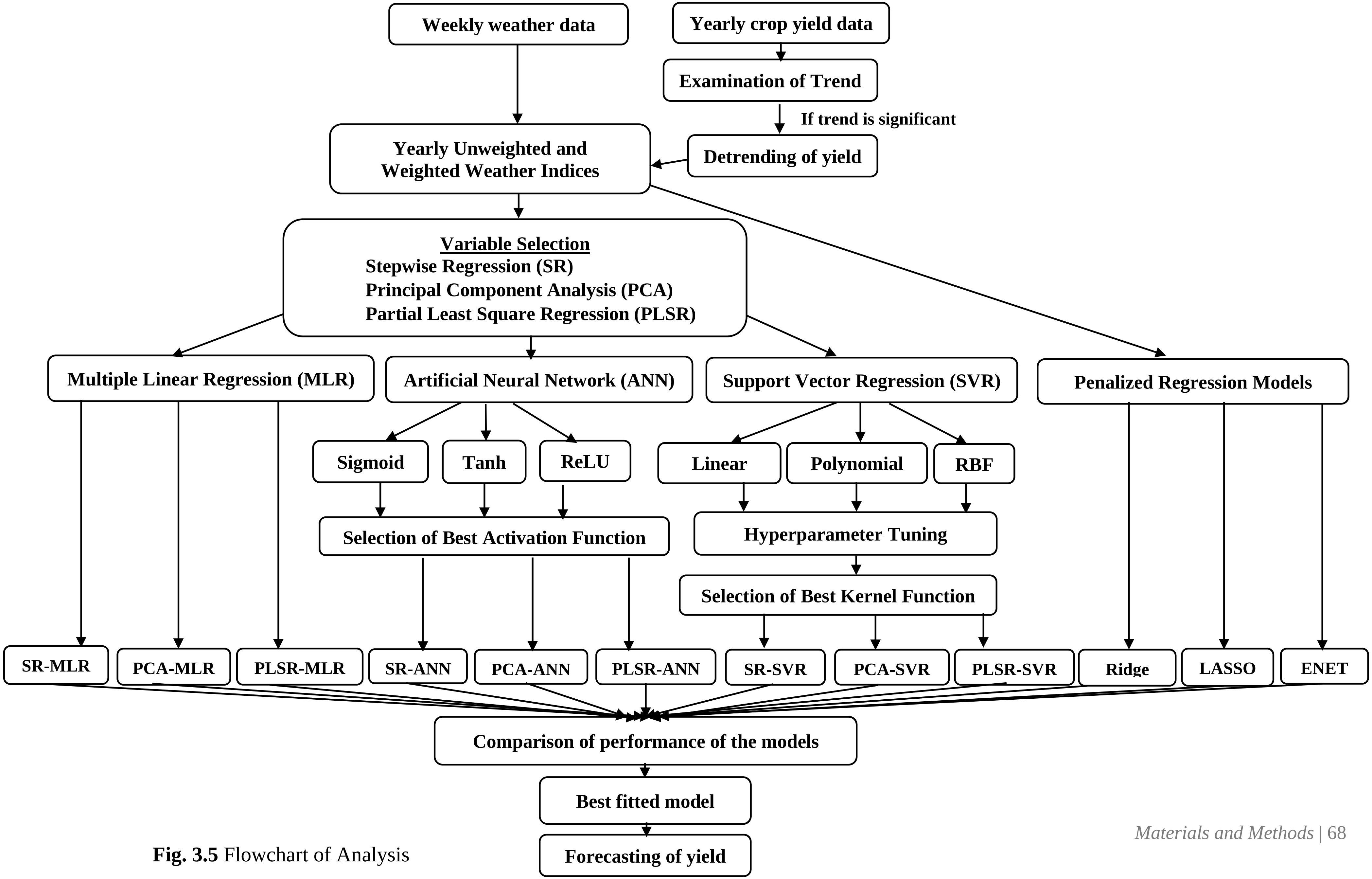

Fig. 3.5 Flowchart of Analysis

Chapter – IV

RESULTS AND DISCUSSION

Chapter IV

RESULTS AND DISCUSSION

In accordance with the objectives formulated for the present study in the introduction chapter, this chapter deals with the presentation and interpretation of the results obtained along with relevant discussions. The results obtained from the present investigation have been divided into the following four sections:

4.1 Calculation of Agrometeorological Indices.

4.2 Calculation of Unweighted and Weighted Weather Indices.

4.3 Application of Variable Selection or Dimension Reduction Methodologies.

4.4 Fitting and Estimation of parameters of different Statistical and Machine Learning Models.

4.5 Comparison of performance of fitted models.

4.6 Forecasting the yield of major crops

4.1 Calculation of Agrometeorological Indices

The agrometeorological indices are the crop-climate related parameter to assess the climate suitability of a location for the cultivation of crops. The agrometeorological indices such as Growing Degree Days (GDD), Diurnal Temperature Range (DTR), Relative Temperature Disparity (RTD) and Number of Rainy Days have been calculated for each crop in each district using the weather data of the period in which the crop is grown. Since, wheat, potato and rapeseed-mustard crops are grown in the same period during the winter season, the agrometeorological indices are calculated commonly for winter crops. For rice and jute crops, the agrometeorological indices are calculated separately for the period in which the crop is cultivated as given in Table 3.1.

The summary of five agrometeorological indices for four districts is given in Table 4.1. The basic idea for calculating the GDD is that the growth and development of the plant will occur only when the temperature exceeds some threshold or base temperature. The GDD of winter crops is high followed by rice and jute season in all the four districts. The higher GDD in the winter season than other two season can be attributed that the threshold temperature requirement of winter crop is low. The GDD of Malda is high for winter crops as well as rice and jute crops followed by Uttar Dinajpur and Jalpaiguri. The GDD of Cooch Behar is comparatively low. There is a significant increasing trend in the GDD of Uttar Dinajpur, Cooch Behar and Malda districts during *aman* rice crop season.

Table 4.1 The summary of Agrometeorological indices in four districts

Indices	District	Winter crops			Rice			Jute		
		Mean	**Sen's slope**	**p value**	**Mean**	**Sen's slope**	**p value**	**Mean**	**Sen's slope**	**p value**
Growing Degree Days (GDD) (º day)	Cooch Behar	3323.59	3.91	0.34	2583.11	5.81	0.02	1609.03	0.55	0.88
	Jalpaiguri	3344.59	-0.67	0.93	2609.04	1.69	0.20	1655.91	0.88	0.54
	Malda	3579.65	-19.84	0.00	2783.30	3.78	0.04	1882.28	-0.48	0.76
	Uttar Dinajpur	3536.09	7.49	0.26	2733.90	9.18	0.02	1815.91	0.92	0.70
Diurnal Temperature Range (DTR) (ºC)	Cooch Behar	14.06	0.08	0.00	8.20	0.06	0.00	9.50	0.00	1.00
	Jalpaiguri	13.02	0.07	0.04	8.40	0.04	0.10	9.99	0.00	0.96
	Malda	11.09	-0.05	0.00	7.05	-0.02	0.16	10.26	-0.05	0.05
	Uttar Dinajpur	12.50	-0.08	0.04	8.56	-0.08	0.02	11.02	-0.05	0.14
Relative Temperature Disparity (RTD) (%)	Cooch Behar	53.00	0.22	0.01	25.58	0.17	0.10	30.10	-0.01	0.81
	Jalpaiguri	49.43	0.21	0.09	25.74	0.01	0.57	30.68	0.01	0.92
	Malda	41.46	-0.02	0.67	21.55	-0.14	0.06	29.53	-0.10	0.06
	Uttar Dinajpur	46.29	-0.43	0.01	25.94	-0.29	0.01	31.90	-0.18	0.06
Rainy Days (RD)	Cooch Behar	3	0.00	0.95	60	-0.67	0.00	37	0.06	0.70
	Jalpaiguri	3	0.00	1.00	68	0.00	1.00	37	0.15	0.19
	Malda	4	-0.33	0.00	47	-0.16	0.43	17	-0.07	0.46
	Uttar Dinajpur	1	0.00	0.31	36	-0.68	0.21	17	-0.27	0.07

The change per year is high in Uttar Dinajpur (9.18 °days) followed by Cooch Behar (5.81 °days) and Malda (3.78 °days). The change in GDD per year in Jalpaiguri is low due to a non-significant trend in GDD. There is a significant decreasing trend in the GDD of Malda during the winter season. The decreasing trend in GDD is attributed to the decreasing trend in the maximum temperature as well as the mean temperature (Mishra, 2013).

The Diurnal Temperature Range (DTR) reflects the variation in temperature within a day. The average DTR is high in the winter season followed by jute and aman rice season. There is an increasing trend in the DTR of Cooch Behar and Jalpaiguri districts and decreasing trend in Malda and Uttar Dinajpur in all three seasons. The DTR is high in Cooch Behar during winter season and it is high in Uttar Dinajpur during jute season.

The Relative Temperature Disparity (RTD) is high during winter season crops followed by jute and *aman* rice season. The higher value of RTD indicates the prevalence of high mean temperature and comparatively lower minimum temperature (Pal *et al* 2013). The RTD is high during the winter season. The Cooch Behar district has registered a high RTD in the winter season followed by Jalpaiguri and Uttar Dinajpur. The RTD of *aman* rice growing season is less. There is a significant increasing trend in the RTD of Cooch Behar and there is a significant negative trend in the RTD of Uttar Dinajpur during winter season. There is a significant decreasing trend in the RTD of Uttar Dinajpur during *aman* rice season.

The number of rainy days is higher during rice season due to the coincidence of monsoon followed by jute season due to summer showers. The number of rainy days during winter is low. It can be observed that there is a significant decreasing trend in the number of rainy days in Malda during winter season and there is a significant decreasing trend in Cooch Behar during *aman* rice season.

4.2 Calculation of Unweighted and Weighted Weather Indices

The weekly weather data of the period in which a particular crop is grown are used to calculate the unweighted and weighted weather indices. The unweighted indices have been calculated as the simple average of each weather variable as given in the formula 3.4. for all the weather parameters except rainfall for which the cumulative value of the crop growing period is calculated instead of average.

4.2.1 Examination of Trend in Yield

Before calculating the correlation coefficient-based as well as path coefficient-based weighted indices, the yield of each crop in each district is tested for the presence of a trend. Modified Mann Kendall is employed to detect the presence of a significant trend in the yield. The null hypothesis of the test is that there is no significant trend in the yield. Simple linear regression is also applied by regressing the yield with year number as the regressor variable. Further, the change in yield per year has been quantified using Sen's slope method. The yield is detrended if the trend in yield is found to be significant.

4.2.1.1 Rice

The results of the trend analysis for rice crop have been given in Table 4.2. The average yield of rice crop is high in Malda (2618.70 kg ha^{-1}) followed by Uttar Dinajpur (2199.57 kg ha^{-1}), Cooch Behar (1966.52 kg ha^{-1}) and Jalpaiguri (1928.15 kg ha^{-1}). From the results of simple regression and the Modified Mann-Kendall Test, it can be seen that there is a highly significant trend in the yield of rice crop in all four districts. Further, the change in the yield of rice crop per year is a high level of 70 kg ha^{-1} per year in Cooch Behar followed by Jalpaiguri district.

Table 4.2 The results of Trend analysis for Rice crop in all four districts

District	Average Yield (kg ha^{-1})	Simple Regression		Modified Mann-Kendall Test		Sen's Slope
		F statistic	p value	Z statistic	p value	
Cooch Behar	1966.52	157.27	0.00	5.20	0.00	70.00
Jalpaiguri	1928.15	254.10	0.00	5.65	0.00	69.95
Malda	2618.70	36.72	0.00	3.85	0.00	62.35
Uttar Dinajpur	2199.57	64.74	0.00	4.59	0.00	62.50

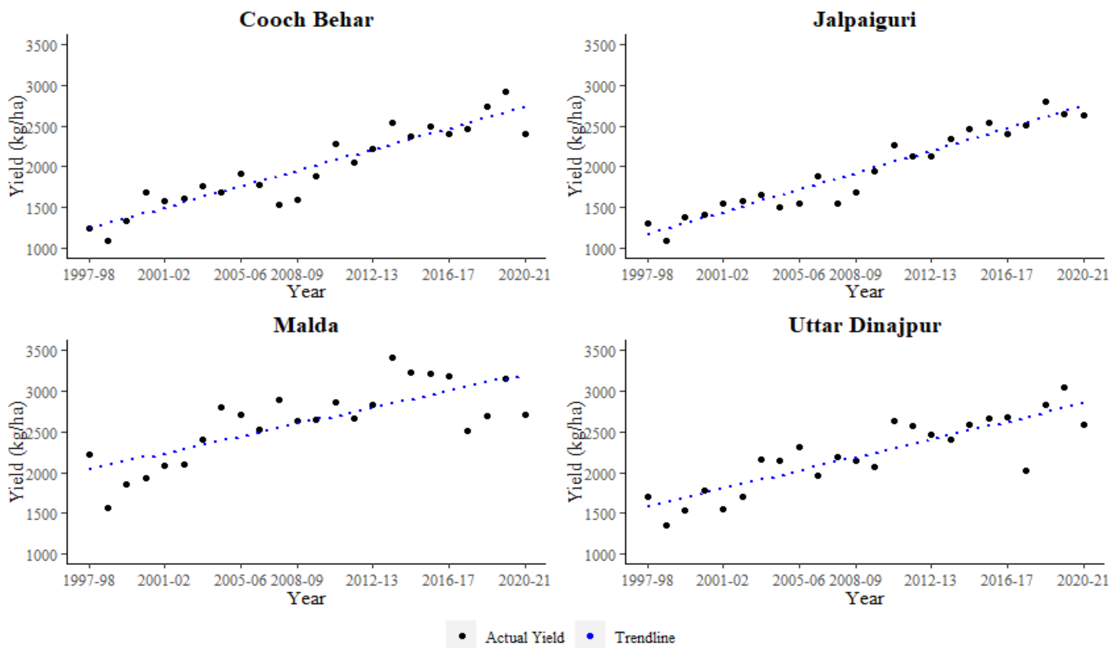

Fig. 4.1 Trend plots for yield of Rice crop in all four districts

It can be seen from Fig. 4.1 that there is an increasing trend in the yield of rice crop in all four districts. The significant increasing trend in the yield of rice crop can be attributed to the introduction high high-yielding varieties, improved cultivational practices *etc.* over the years. As there is a significant trend in the yield of rice crop in all the four districts, the yield is to be determined before the calculation of the correlation coefficient as well as path coefficient-based weighted indices.

4.2.1.2 Wheat

The results of the trend analysis for wheat crop have been given in Table 4.3. The average yield of wheat crop is high in Malda (2749.32 kg ha^{-1}) followed by Uttar Dinajpur (2382.41 kg ha^{-1}), Jalpaiguri (2136.43 kg ha^{-1}) and Cooch Behar (2022.81 kg ha^{-1}). It can be observed from the results of trend analysis that there is a highly significant trend at 0.01% level of significance in the yield of wheat crop in all the districts except Uttar Dinajpur. The trend in the yield of wheat crop in Uttar Dinajpur is also significant at 5% level of significance as the p value is less than 0.05. Further, the change in the yield of wheat crop per year is a high level of 37.62 kg ha^{-1} per year in Cooch Behar followed by Jalpaiguri.

Table 4.3 The results of Trend analysis for Wheat crop in all four districts

District	Average Yield (kg ha^{-1})	Simple Regression		Modified Mann-Kendall Test		Sen's Slope
		F statistic	p value	Z statistic	p value	
Cooch Behar	2022.81	10.69	0.00	2.85	0.00	37.62
Jalpaiguri	2136.43	10.38	0.00	7.68	0.00	36.86
Malda	2749.32	22.91	0.00	3.64	0.00	33.34
Uttar Dinajpur	2382.41	6.92	0.02	2.11	0.03	14.79

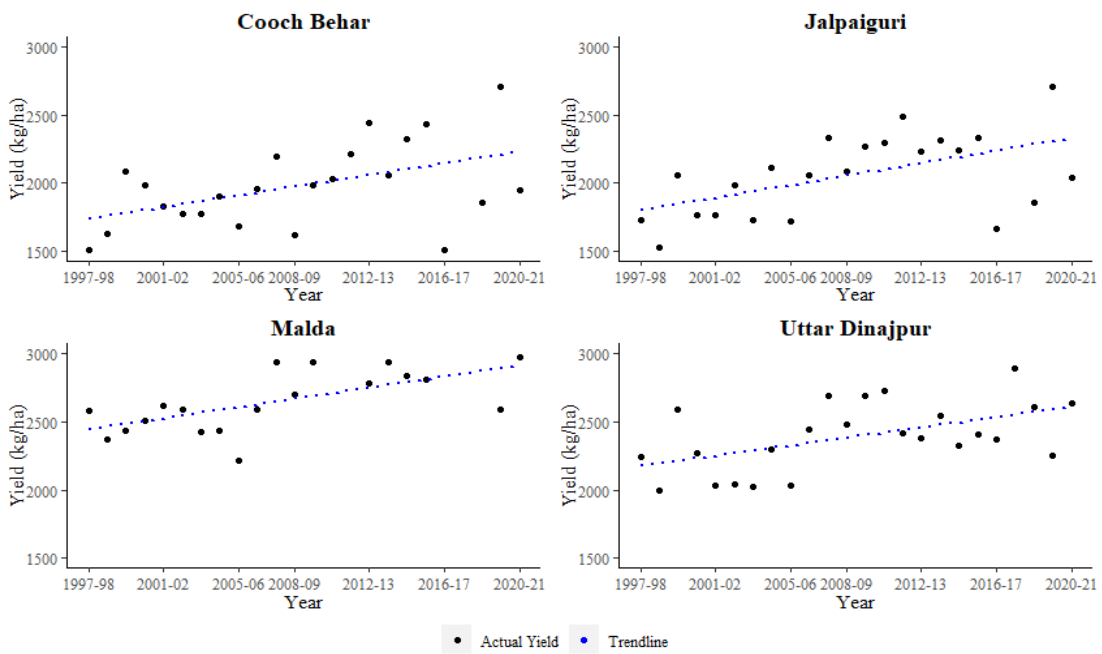

Fig. 4.2 Trend plots for yield of Wheat crop in all four districts

It can be seen from Fig. 4.2 that there is an increasing trend in the yield of wheat crop in all four districts.

4.2.1.3 Potato

The results of the trend analysis for potato crop have been given in Table 4.4. The average yield of potato crop is high in Malda (26.52 t ha^{-1}) followed by Jalpaiguri (24.86 t ha^{-1}), Cooch Behar (23.97 t ha^{-1}) and Uttar Dinajpur (23.81 t ha^{-1}). There is a highly significant trend at 0.01% level of significance in the yield of potato crop in all the districts. Further, the change in the yield of potato crop per year is a high level of 1.17 t ha^{-1} per year in Malda.

Table 4.4 The results of Trend analysis for Potato crop in all four districts

District	Average Yield	Simple Regression		Modified Mann-Kendall Test		Sen's Slope
	(t ha^{-1})	F statistic	p value	Z statistic	p value	
Cooch Behar	23.97	8.05	0.00	3.11	0.00	0.55
Jalpaiguri	24.86	35.86	0.00	4.12	0.00	0.72
Malda	26.52	29.49	0.00	3.53	0.00	1.17
Uttar Dinajpur	23.81	33.21	0.00	4.06	0.00	0.65

It can be seen from Fig. 4.3 that there is an increasing trend in the yield of potato crop in all four districts.

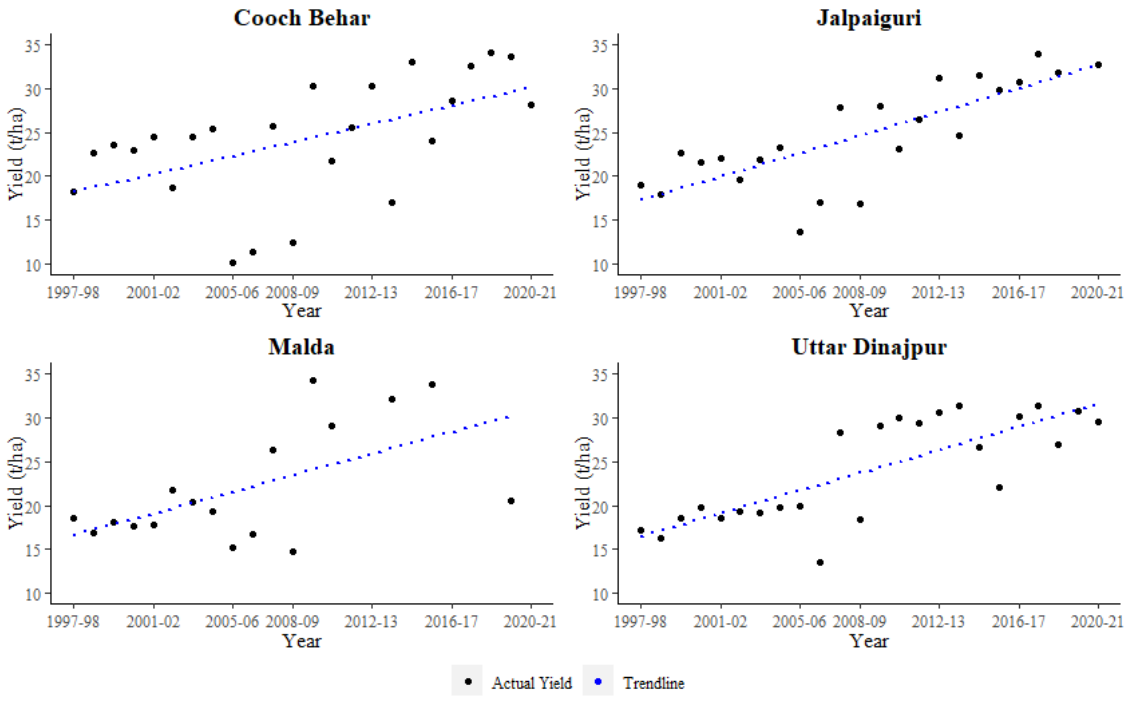

Fig. 4.3 Trend plots for yield of Potato crop in all four districts

4.2.1.4 Jute

The results of trend analysis for jute crop have been given in Table 4.5. The average yield of jute crop is high in Malda (13.96 bales ha^{-1}) and the average yield in other districts are less than 12 bales ha^{-1}. There is a highly significant trend at 0.01% level of significance in the yield of jute crop in all the districts. Further, the change in the yield of jute crop per year is high in Malda (0.27 bales ha^{-1} per year) followed by Jalpaiguri and Uttar Dinajpur districts.

Table 4.5 The results of Trend analysis for Jute crop in all four districts

District	Average Yield (bales ha^{-1})	Simple Regression		Modified Mann-Kendall Test		Sen's Slope
		F statistic	p value	Z statistic	p value	
Cooch Behar	11.25	78.38	0.00	4.85	0.00	0.23
Jalpaiguri	11.80	86.51	0.00	4.59	0.00	0.28
Malda	13.96	58.72	0.00	4.27	0.00	0.31
Uttar Dinajpur	11.78	71.46	0.00	5.36	0.00	0.27

There is an increasing trend in the yield of jute crop in all four districts as seen from Fig. 4.4.

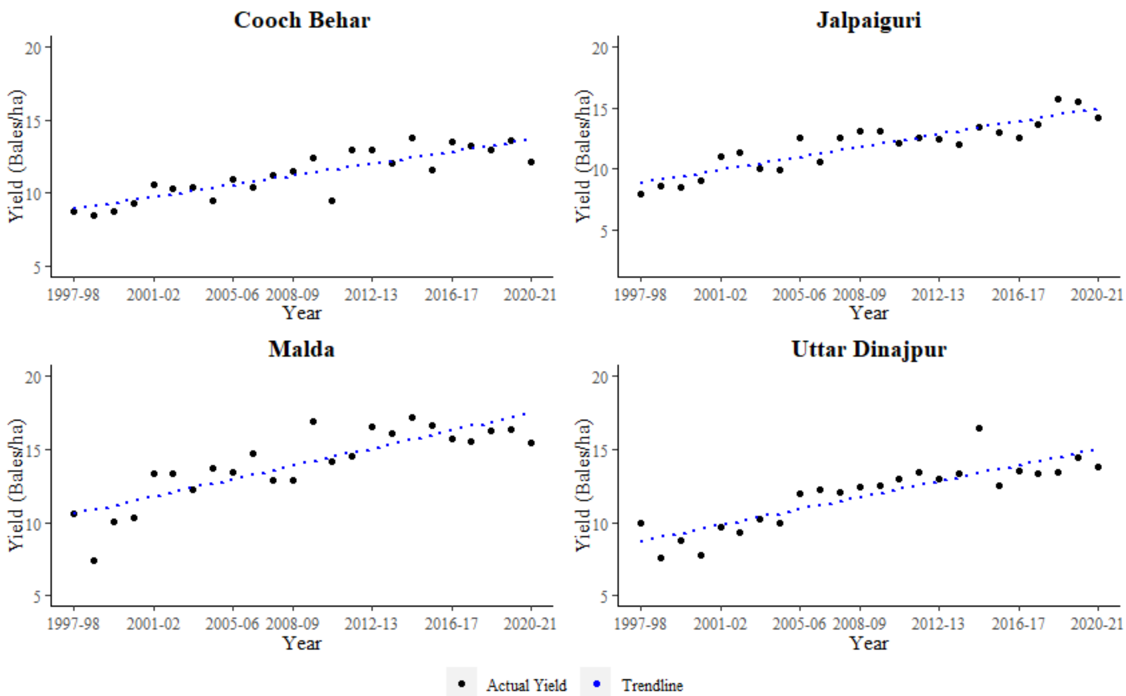

Fig. 4.4 Trend plots for yield of Jute crop in all four districts

4.2.1.5 Rapeseed-Mustard

The results of the trend analysis for rapeseed-mustard crop have been given in Table 4.6. The average yield of rapeseed-mustard crop in Malda is a high level of 1054.54 kg ha^{-1} and the average yield is less than 800 kg ha^{-1} in the remaining districts. From the results of simple regression as well as the Modified Mann-Kendall Test, it can be revealed that there is a highly significant trend in the yield of rapeseed-mustard crop in Jalpaiguri, Malda and Uttar Dinajpur districts which are significant at 0.01% level of significance. But there is no significant trend in Cooch Behar at 5% level of significant (p>0.05). The change per year in the yield is high in Uttar Dinajpur (22.49 kg ha^{-1}) followed by Malda (21.92 kg ha^{-1}) and Jalpaiguri (13.41 kg ha^{-1}). The change is very low (7.29 kg ha^{-1}) in Cooch Behar which might be due to the non-significant trend in yield.

Table 4.6 The results of Trend analysis for Rapeseed-mustard crop in all four districts

District	Average Yield (kg ha^{-1})	Simple Regression		Modified Mann-Kendall Test		Sen's Slope
		F statistic	p value	Z statistic	p value	
Cooch Behar	539.06	2.76	0.11	1.16	0.24	7.29
Jalpaiguri	626.17	31.57	0.00	4.33	0.00	13.41
Malda	1054.54	35.46	0.00	6.85	0.00	21.92
Uttar Dinajpur	799.31	33.41	0.00	3.85	0.00	22.49

Due to the non-significant trend in the yield of rapeseed-mustard in Cooch Behar district, the slope of the trendline is very low as seen from Fig. 4.5. There is an increasing trend in the yield of rapeseed-mustard crop in the other three districts.

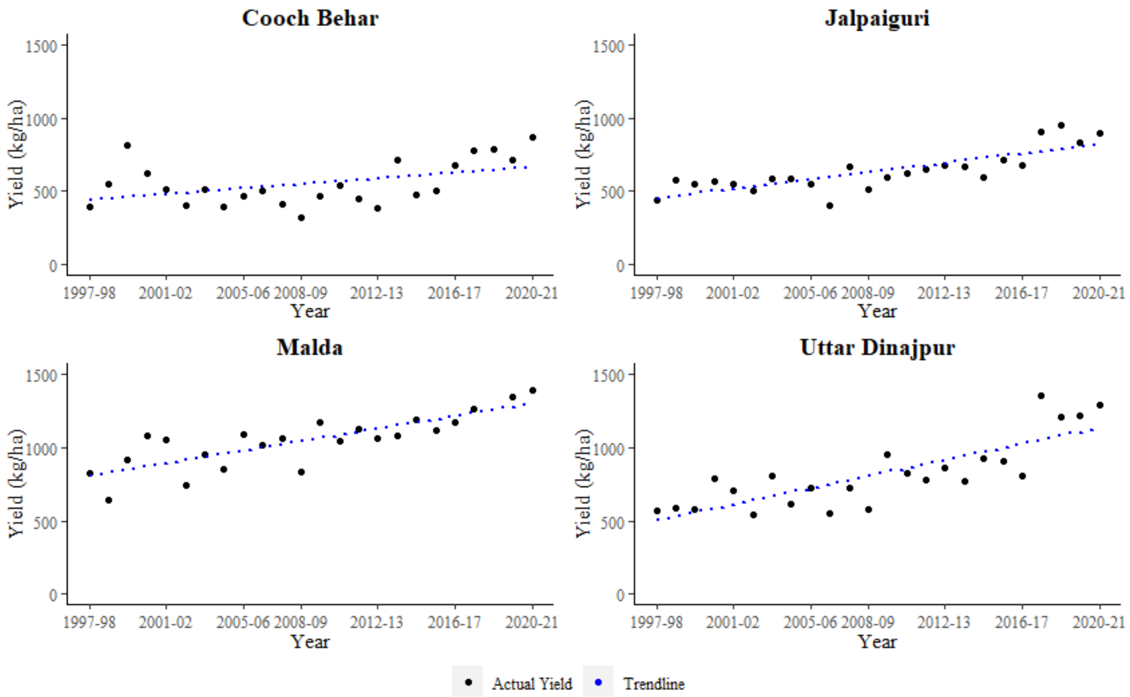

Fig. 4.5 Trend plots for yield of Rapeseed-mustard crop in all four districts

Since there is a significant trend in the yield of all the five crops in all four districts except rapeseed-mustard yield in Cooch Behar, the respective yield is detrended before the calculation of weighted weather indices. Detrending of yield before calculating correction as well as path coefficient is done with an aim to study the actual effect of weather parameters on yield by removing trend-causing factors. The detrending of yield has been done by subtracting the predicted yield from actual yield where the predicted yield is obtained by regressing the yield using year number as the explanatory variable.

4.2.2 Calculation of Weighted Indices

In order to calculate the correlation coefficient based on weighted indices, first-week data of a particular weather parameter say, the maximum temperature of all the years have been taken and the correlation coefficient between the first-week data and the detrend yield is calculated. Similarly, second-week data of all the years are taken and the correlation between the second-week data and the detrend yield is calculated. This procedure is continued till the last week of the crop growing period of a particular crop. Then, the correlation coefficient-based weighted indices have been calculated for each weather parameter using the formula 3.5 as the weighted average of weekly weather data with the correlation coefficient as the weight. In similar fashion, the path coefficient-based weighted indices have been calculated as the weighted average of weather variables as given in the formula 3.6. The average of unweighted and correlation coefficient based as well as path coefficient based weighted indices of all the five crops in each district are given in the Table 4.7.

In rice crop, the correlation coefficient based weighted indices of maximum temperature (CC_Tmin) of all the district are almost close to unweighted indices except Uttar Dinajpur, where the CC_Tmax is greater than unweighted based on the correlation of maximum temperature with yield. CC_RF of all the districts is far lesser than the unweighted according to the effect of rainfall on yield. CC_WS of Uttar Dinajpur is negatively weighted that might be due to the negative effect of windspeed on yield. The PC_Tmin of Cooch Behar is weighted more according to its direct effect on yield; whereas, it is negatively weighted in Jalpaiguri. The exposure of different phenological phases of rice crop to a temperature lower than the threshold temperature or a cold spell affects the yield directly (Nishad *et al* 2018). The PC_RF of all the districts is also far lesser than unweighted indices but it is close to the CC_RF. Undistributed rainfall during the monsoon season largely influences the kharif rice cop which ultimately affects the yield (Akhter *et al* 2021). The correlation coefficient as well as path coefficient-based weighted indices of relative humidity are also closer to unweighted indices.

For the wheat crop in Cooch Behar, the average of correlation-based indices of all the weather parameters are lesser than the unweighted indices. The reduction in the weighted indices may be according to the correlation of weather parameters with yield in different weeks. The correlation coefficient-based weighted index of maximum temperature (CC_Tmax) of the other three districts is greater than the unweighted indices which might be due to the importance of maximum temperature to the yield of wheat crop. CC_Tmin as well as CC_RF are also greater than unweighted indices in the case of Malda district.

Table 4.7 Average of Unweighted and Weighted Weather Indices of five crops in four districts

Crop	Rice				Wheat				Potato				Jute				Rapeseed-mustard			
Indices	**Cooch Behar**	**Jalpaiguri**	**Malda**	**Uttar Dinajpur**	**Cooch Behar**	**Jalpaiguri**	**Malda**	**Uttar Dinajpur**	**Cooch Behar**	**Jalpaiguri**	**Malda**	**Uttar Dinajpur**	**Cooch Behar**	**Jalpaiguri**	**Malda**	**Uttar Dinajpur**	**Cooch Behar**	**Jalpaiguri**	**Malda**	**Uttar Dinajpur**
Tmax (ºc)	31.6	31.9	32.4	32.6	26.5	26.2	26.9	31.9	26.5	26.2	26.9	27.2	31.4	32.1	34.7	34.3	26.5	26.2	26.9	27.2
Tmin (ºc)	23.4	23.5	25.3	24.0	12.5	13.3	15.8	19.0	12.5	13.3	15.8	14.7	21.9	22.2	24.3	23.2	12.5	13.3	15.8	14.7
RH (%)	82.4	82.0	80.0	82.3	75.5	76.2	71.2	77.4	75.5	76.2	71.2	80.1	74.5	73.6	67.0	78.7	75.5	76.2	71.2	80.1
WS (m/s)	2.1	2.2	2.8	2.5	1.8	1.9	2.6	3.2	1.8	1.9	2.6	2.3	2.5	2.7	3.1	3.0	1.8	1.9	2.6	2.3
RF (mm)	2060	2380	1004	899	45.1	32.3	67.0	5.7	45.1	32.3	67.0	15.3	1071	983	354	335	45.1	32.3	67.0	15.3
CC_Tmax	29.4	30.2	32.1	38.4	25.0	28.3	31.1	34.2	24.9	24.1	38.5	20.3	31.5	23.3	34.2	34.0	22.6	25.4	22.4	26.3
CC_Tmin	26.0	2.3	24.6	23.4	9.5	2.1	32.4	15.4	3.5	12.4	38.8	-3.5	21.7	21.5	45.8	22.2	11.9	12.7	13.7	15.6
CC_RH	83.3	80.3	77.9	82.4	73.5	74.9	62.3	79.9	82.6	79.1	49.0	80.6	65.9	77.8	56.6	79.3	37.4	87.7	-217.7	83.1
CC_WS	2.1	3.9	2.2	-2.9	0.6	2.7	2.6	2.3	2.0	1.9	2.6	2.1	2.4	2.8	3.3	3.9	1.5	1.8	2.6	2.4
CC_RF	123.2	115.1	39.3	73.8	22.5	0.8	5.7	3.2	-97.5	-2.4	4.4	-0.3	135.8	127.6	37.5	61.0	10.6	14.1	-1.4	1.3
PC_Tmax	32.1	31.3	32.2	30.7	25.4	22.6	25.6	30.2	29.4	27.6	23.9	25.4	31.6	36.0	33.6	34.3	25.5	26.3	23.6	27.6
PC_Tmin	78.0	-94.7	25.9	24.0	2.8	-227.3	5.5	15.9	18.1	13.0	-41.9	30.5	22.2	23.0	31.1	20.8	11.9	13.7	14.0	12.4
PC_RH	83.0	80.7	78.9	82.3	107.7	76.5	111.8	80.5	80.9	87.2	190.2	80.6	74.7	72.9	53.9	79.2	69.3	7.2	74.0	80.3
PC_WS	2.1	2.3	-2.3	2.8	1.4	2.4	2.5	2.2	2.3	1.8	2.6	2.1	2.5	3.1	3.0	3.5	1.7	2.1	2.7	2.4
PC_RF	112.5	62.5	48.1	76.7	2.1	-0.9	5.2	5.9	-17.5	2.1	4.5	-1.2	144.5	137.4	54.9	33.4	1.6	2.3	3.1	-0.4

Similarly, the path coefficient weighted indices are also optimally weighted for each weather parameter according to its direct effect on yield. Noticeability, PC_RF of all the districts is far lesser than the unweighted indices which might be due to their lesser direct effect on yield. The PC_RF of Jalpaiguri is negative which might be due to the negative effect of rainfall on yield. The PC_RH of all the districts is greater than the unweighted indices due to the higher direct effect of relative humidity on wheat yield. The high relative humidity during the 50% flowering stage of the wheat crop is having a direct positive effect on attaining the physiological maturity of the crop (Satish *et al* 2023). The PC_Tmin of Jalpaiguri has a high negative value. The maximum reduction in wheat yield in Jalpaiguri district has been noticed due to the decline of 2.4ºC to 2.8ºC in minimum temperature during winter season (Mukherjee *et al* 2023).

In the case of potato crop also, almost all the correlation coefficient indices are lesser than the unweighted indices except, CC_Tmax and CC_Tmin in Malda. The CC_RF is negative in Cooch Behar, Jalpaiguri and Uttar Dinajpur districts which might be due to the negative effect of rainfall on yield. The path coefficient-based index of maximum temperature (PC_Tmax) are greater than unweighted indices in Cooch Behar, Jalpaiguri and Uttar Dinajpur. The PC_Tmin of Cooch Behar and Uttar Dinajpur are also greater than unweighted indices. The prevalence high maximum and minimum temperature during mid of the winter season has high positive effect on potato tuber yield (Singh and Sandhu, 2022). The PC_Tmin of Malda is negative which may due to the negative direct effect of minimum temperature on potato yield. The PC_RH of all the districts is greater than the unweighted indices which might be due to the greater direct effect of relative humidity as similar to the wheat crop. The relative humidity of greater than 85% increase the tuber yield significantly (Wheeler *et al* 1989). The PC_RF is negative in Cooch Behar and Uttar Dinajpur districts and it is positive but lesser than the unweighted indices in Jalpaiguri and Malda districts.

The CC_Tmax is less weighted than unweighted indices in Jalpaiguri and CC_Tmin is more weighted in Malda according to the effect of the respective weather parameter on yield of jute crop. The correlation coefficient-based weighted indices of windspeed and relative humidity are almost close to the unweighted indices of respective weather factors. CC_RF of all the districts is far lesser than the unweighted according to the effect of rainfall on yield. The PC_Tmax of Jalpaiguri and CC_Tmin of Malda are more weighted which might be due to greater direct effect of the respective weather parameter on yield. An increased temperature of 0.8ºC to1.4ºC during jute growing season increases the fibre yield (Singh *et al* 2019). PC_RH of Malda are lesser than the unweighted as well as correlation-based weighted indices according to their direct effect on yield. Similar to CC_RF, the PC_RF of all the districts is far lesser than the unweighted indices and there is a considerable difference between PC_RF and CC_RF.

Similarly, correlation coefficient indices are lesser than the unweighted indices of rapeseed-mustard crop except for CC_RH of Jalpaiguri and Uttar Dinajpur districts. The CC_RH and CC_RF of Malda are negative which might be due to the negative association of relative humidity and rainfall with yield. The path coefficient-based indices of almost all the weather parameter are lesser or closer to unweighted indices. The PC_RF is far lesser than unweighted as well as correlation-based weighted indices that might be due to the lesser direct effect of rainfall on rapeseed-mustard yield. Intermitted rainfall during winter season affects the Indian mustard yield considerably (Boomiraj *et al* 2010).

It can be observed that both the weighted indices of majority of the weather parameters are lesser than weighted indices in most of the cases and some of the weighted indices are negative due to the negative association of the weather parameter on yield. Hence, the weighted indices differ from unweighted indices in terms of weighting a particular weather parameter based on its relationship on yield. The correlation-based weighted indices are weighting a particular weather parameter according to its correlation with yield, while path coefficient-based weighted indices weighting a particular weather parameter according to its direct effect on yield. Whereas, the unweighted indices are just an average of the weather parameter during the weeks in which the particular crop is grown.

4.3 Application of Variable Selection or Dimension Reduction Methodologies

Five unweighted, five correlation coefficients based and five path coefficient based weighted indices have been calculated for each crop in each district as there are five weather parameters. Since the trend in yield is significant in almost all crops in every district, the year number is also used as a trend variable in the model to account the trend-causing factor along with weather indices. It is to be mentioned here that the detrended yield is only used while calculating the weather indices with an aim of studying the exact effect of weather factors on yield by removing the trend. However, the actual yield is used while fitting the model. Hence, the weather indices as well as time trend variable are used as the input or explanatory variables. The time trend variable is referred as Year hereafter. Therefore, there are sixteen input or explanatory variables for each crop in each district.

The inclusion of all the explanatory variables in the model leads to a complex model which necessitates the estimation of many parameter values that leads to the slow learning of the model. In some cases, inclusion of some of the redundant or nonsignificant variables into the model leads to the overfitting issue. With this aim, three variable selection or dimension reduction methodologies *viz.* Stepwise Regression (SR), Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR) are employed.

4.3.1 Stepwise Regression (SR)

The stepwise regression analysis has been carried out by taking crop yield as the dependent variable and fifteen weather indices as well as time trend variables as explanatory variables for each district separately.

4.3.1.1 Rice

The summary of stepwise regression analysis for rice crop in all four districts is given in Table 4.8. For the Cooch Behar district, the time trend variable is included in the first step which is significant at 1% Level of Significant (LoS). The Adjusted R^2 is 0.86 and the Akaike's Information Criteria (AIC) value is 306.70. The CC_RF is included in the second step which further increases the adjusted R^2 to 0.93. There is a decrease in AIC value as well. CC_WS, PC_RF, Tmin and CC_Tmax are included in subsequent steps. There is a further increase in adjusted R^2 and a decrease in AIC values on the inclusion of the above variables. All six indices and intercepts are significant at 5% LoS. These six indices cumulatively explain 98% of the variation in the yield. The AIC value is at a low level of 262.54 when these seven indices are included in the model. The inclusion of other indices to the stepwise regression does not increase the adjusted R^2 to a significant level and their respective regression coefficients are not significant at 5% LoS.

There are five steps for Jalpaiguri district in which Trend variable, CC_RF PC_WS, PC_RH and PC_RF are included in the model in each step respectively. There is a decline in the AIC value in each step and it is low when these five indices are included in the model. These five indices are significant at 5% LoS. These five indices are together explain 98% of the variation present in the yield of rice crop in Jalpaiguri district. Similarly, there are five steps for Malda district in which trend variable, CC_WS, CC_RH, CC_RF and PC_Tmax are included in each step respectively. The time trend variable, CC_Tmax and PC_RF are included in stepwise regression for Malda district. All the five indices and intercept are significant at 5% LoS. The AIC value is decreased and Adjusted R^2 is increased on inclusion of each variable into the model. The AIC value is low when these five indices are included in the model. These five indices cumulatively explain 92% of the variation in the yield of rice crop in Malda district.

The trend variable, CC_Tmin and CC_RF are included in three steps for Uttar Dinajpur district. All three indices along with intercept are significant at 1% LoS and

AIC is low when these three indices are included in the model. These three indices cumulatively explain 93% of the variation in the yield of rice crop in Uttar Dinajpur district.

Table 4.8 Summary of Stepwise Regression (SR) analysis for Rice crop

District	Number of Steps	Predictors	Coefficient	p value
Cooch Behar	6	(Intercept)	2152.33	0.00
		Year	71.00	0.00
		CC_RF	-0.26	0.01
		CC_WS	-549.80	0.00
		PC_RF	-0.88	0.00
		Tmin	70.71	0.00
		CC_Tmax	-13.10	0.02
		CC_Tmin	-39.19	0.05
Jalpaiguri	5	(Intercept)	96.12	0.84
		Year	69.18	0.00
		CC_RF	-0.19	0.01
		PC_WS	-176.08	0.00
		PC_RH	17.80	0.01
		PC_RF	-0.07	0.05
Malda	5	(Intercept)	7160.48	0.00
		Year	54.98	0.00
		CC_WS	79.77	0.02
		CC_RH	-24.21	0.01
		CC_RF	-2.70	0.02
		PC_Tmax	-105.12	0.04
Uttar Dinajpur	3	(Intercept)	511.39	0.00
		Year	63.47	0.00
		CC_Tmin	36.09	0.00
		CC_RF	1.12	0.00

4.3.1.2 Wheat

The results of stepwise regression analysis for wheat crop for all four districts is given in Table 4.9. For the Cooch Behar district, CC_Tmin is included in the first step which is significant at 1% LoS. CC_Tmin is alone explains 60% variation in the yield. The RF is included in the second step which further increases the adjusted R^2 to 0.70. There is a decline in AIC value as well. Tmin is included in the third step. There is a further increase in adjusted R^2 and a decrease in AIC values. All three includes indices and intercept are significant at 1% LoS. These three indices cumulatively explaining 73% of the variation in the yield of wheat crop. The inclusion of other indices to the stepwise regression does not increase the adjusted R^2 to a significant level and their respective regression coefficients are not significant at 5% LoS.

There are four steps for the Jalpaiguri district in which PC_Tmin, RH, Time trend and CC_RF are added in successive steps. There is a decline in the AIC value in each step and it is low when these four indices are included in the model. These four indices are together explain 88% of the variation present in the yield of wheat crop in Jalpaiguri district. These four indices are significant at 1% LoS. The time trend variable, CC_Tmax and PC_RF are included in stepwise regression for the Malda district. All three included indices are significant at 1% LoS. AIC value is low when these three indices are included in the model. These three indices cumulatively explain 83% of the variation in the yield of wheat crop in Malda district.

The CC_Tmin, CC_RF and RF are included in three steps for Uttar Dinajpur district. The AIC value is 312.81 and the adjusted R^2 is 0.38 in the first step in which CC_Tmin alone is included in the model. There is a decrease in AIC value and an increase in adjusted R^2 in the subsequent steps. These three indices cumulatively explain 73% of the variation in the yield of wheat crop in Uttar Dinajpur district. All three indices along with intercept are significant at 1% LoS.

Table 4.9 Summary of Stepwise Regression (SR) analysis for Wheat crop

District	Number of Steps	Predictors	Coefficient	p value
Cooch Behar	3	(Intercept)	3092.47	0.00
		CC_Tmin	-208.08	0.00
		RF	-3.15	0.00
		Tmin	84.36	0.00
Jalpaiguri	4	(Intercept)	-1147.93	0.31
		PC_Tmin	1.08	0.00
		RH	41.52	0.01
		Year	31.49	0.00
		CC_RF	-15.72	0.00
Malda	3	(Intercept)	147.72	0.69
		Year	37.73	0.00
		CC_Tmax	67.42	0.00
		PC_RF	10.52	0.00
Uttar Dinajpur	3	(Intercept)	903.64	0.00
		CC_Tmin	101.96	0.00
		CC_RF	-43.42	0.00
		RF	8.66	0.01

4.3.1.3 Potato

The results of stepwise regression analysis for potato crop for all four district is given in the Table 4.9. For Cooch Behar district, CC_Tmin is included in the first step which is significant at 5% LoS. The Adjusted R^2 is 0.32 when CC_Tmin alone included in the model. The AIC value is 149.63. The CC_WS is included in the second step which further increases the adjusted R^2 to 0.56. There is a decline in AIC value as well. CC_RF is included in the third step. There was a further increase in adjusted R^2 and a decrease in AIC values. The time trend variable is included in the fourth step. All the four included indices are significant at 5% LoS. These three indices cumulatively explain 66% of the variation in the yield of potato crop.

There are five steps for the Jalpaiguri district in which CC_Tmin, Year, PC_RH, PC_Tmax and PC_WS are added in successive steps. The AIC is 130.46 and the adjusted R^2 is 0.62 when CC_Tmin alone is included in the model in first step. There is a decline in the AIC value in each step and it is low when these five indices are included in the model. These five indices explain 92% of the variation present in the yield of potato crop in Jalpaiguri district. These five indices and intercept are significant at 5% LoS. Similarly, there are five steps for Malda district in which the time trend variable, CC_Tmin, PC_RH, RF and PC_Tmax are included in each step respectively. All the five includes indices as well as the intercept are significant at 5% LoS. AIC value is decreasing in each step and the low AIC of 130.70 is obtained in the fifth step. These five indices cumulatively explain 86% of the variation in the yield of potato crop in the Malda district.

Table 4.10 Summary of Stepwise Regression (SR) analysis for Potato crop

District	Number of Steps	Predictors	Coefficient	p value
Cooch Behar	4	(Intercept)	-4.55	0.53
		CC_Tmin	-0.49	0.03
		CC_WS	13.00	0.00
		CC_RF	-0.01	0.04
		Year	0.29	0.05
Jalpaiguri	5	(Intercept)	45.80	0.00
		CC_Tmin	-2.84	0.00
		Year	0.52	0.00
		PC_RH	-0.13	0.00
		PC_Tmax	0.51	0.04
		PC_WS	3.16	0.05
Malda	5	(Intercept)	-40.67	0.00
		Year	1.50	0.00
		CC_Tmin	0.43	0.00
		PC_RH	0.05	0.00
		RF	0.04	0.01
		PC_Tmax	0.83	0.03
Uttar Dinajpur	4	(Intercept)	0.30	0.96
		Year	0.68	0.00
		CC_Tmin	-0.23	0.00
		CC_RF	0.91	0.01
		CC_RH	0.18	0.03

The time trend variable, CC_Tmin, CC_RF and CC_RH are included in four steps for Uttar Dinajpur district. The AIC is low when these three indices are included in the model and all three indices are significant at 5% LoS. These three indices cumulatively explain 88% of the variation in the yield of potato crop in Uttar Dinajpur district.

4.3.1.4 Jute

The results of stepwise regression analysis for jute crop in all four districts is given in Table 4.11. The time trend variable is included in the first step for Cooch Behar district which is significant at 1% LoS. The AIC value is 59.59 and the Adjusted R^2 is 0.78 when the time trend variable alone is included in the model. CC_Tmin, PC_Tmax, CC_WS and PC_RH are included in the subsequent steps. There is a decrease in AIC value as well as an increase in adjusted R^2 in each step. All the five indices included are significant at 5% LoS. These five indices are cumulatively explaining 94% of the variation in the yield of jute crop in Cooch Behar district. The inclusion of other indices to the stepwise regression does not increase the adjusted R^2 to a significant level and their respective regression coefficients are not significant at 5% LoS.

There are four steps for the Jalpaiguri district in which time trend variable, CC_WS, PC_RF and RF are included in the model in successive steps. There is a decline in the AIC value in each step and it is low level of 36.58 when all the four indices are included in the model. These four indices explain 95% of the variation in the yield of jute crop in Jalpaiguri district. These four indices and intercept are significant at 5% LoS. There are three steps for Malda district in which the time trend variable, CC_WS and CC_RF are included in each step respectively. The adjusted R^2 is 0.72 when time trend variable alone is included in the first step. The adjusted R^2 increases in each step and the AIC value decreases in each step. The low AIC value of 66.99 is obtained in the third step. All three indices as well as the intercept are significant at 5% LoS. and. These three indices cumulatively explain 87% of the variation in the yield of jute crop in Malda district.

There are four steps for Uttar Dinajpur district in which the time trend variable, CC_WS, PC_Tmax and CC_RF are included in the stepwise regression model. The adjusted R^2 is 0.76 and AIC value is 73.02 when the time trend variable alone is included in the first step. There is a decline in AIC and an increment in adjusted R^2 in each step. The adjusted R^2 is 0.96 and the AIC value is 34.04 in the fourth step in which four indices are included in the model. All the four indices are significant at 5% LoS. These four indices are cumulatively explained 96% of the variation in the yield of jute crop in Uttar Dinajpur district.

Table 4.11 Summary of Stepwise Regression (SR) analysis for Jute crop

District	Number of Steps	Predictors	Coefficient	p value
Cooch Behar	5	(Intercept)	3.24	0.35
		Year	0.24	0.00
		CC_Tmin	-0.39	0.00
		PC_Tmax	0.44	0.00
		CC_WS	1.10	0.00
		PC_RH	-0.04	0.05
Jalpaiguri	4	(Intercept)	11.77	0.00
		Year	0.31	0.00
		CC_WS	-1.39	0.00
		PC_RF	-0.01	0.00
		RF	0.00	0.04
Malda	3	(Intercept)	13.92	0.00
		Year	0.26	0.00
		CC_WS	-0.80	0.00
		CC_RF	-0.01	0.05
Uttar Dinajpur	4	(Intercept)	2.53	0.25
		Year	0.25	0.00
		CC_WS	-0.35	0.00
		PC_Tmax	0.21	0.00
		CC_RF	0.00	0.04

4.3.1.5 Rapeseed-Mustard

The summary of the stepwise regression analysis for rapeseed-mustard crop is given in the Table 4.12. The CC_RH is included in the first step for Cooch Behar district which is significant at 5% LoS. The AIC value is 282.45 and Adjusted R^2 is 0.50 when CC_RH alone included in the model. PC_Tmax and CC_WS are included in second and third steps. There is a decrease in AIC value as well as increase in adjusted R^2 in each step. All the three indices and intercept are significant at 5% LoS. These three indices cumulatively explain 72% of variation in the yield of rapeseed-mustard crop. The time trend variable is not included in the model as there is no significant trend in yield of rapeseed-mustard in Cooch Behar district.

There are five steps for Jalpaiguri district in which time trend variable, CC_Tmax, PC_WS, Tmin and CC_WS are included to the model in successive steps. The adjusted R^2 is 0.58 and AIC is 274.03 in the first step in which the time trend variable alone included in the model. There is a decline in the AIC value in each step and it is low level of 250.68 when all the five indices are included in the model. These four indices are explaining 87% of the variation in the yield of rapeseed-mustard crop in Jalpaiguri district. These five indices are significant at 5% LoS. There are four steps for Malda district in which the time trend variable, CC_RF, CC_RH and PC_RH are included in each step respectively. The adjusted R^2 is 0.61 when time trend variable alone is included in the first step. The adjusted R^2 increases in each step and AIC value decreases in each step. The low AIC value of 259.25 is obtained in the fourth step after including the above-mentioned indices. All four indices as well as the intercept are significant at 5% LoS. and. These four indices are cumulatively explain 91% of variation in the yield of rapeseed-mustard crop in the Malda district.

There are three steps for Uttar Dinajpur district in which the time trend variable, CC_RH and CC_WS are included the stepwise regression model. The adjusted R^2 is 0.60 and AIC value is 279.08 when time trend variable alone included in the first step. There is a decline in AIC and an increment in adjusted R^2 in each step. The adjusted R^2 is 0.79 and AIC value is 284.15 in the third step in which the two indices and a time trend variable are included in the model. All the three are significant at 1% LoS. These three indices are cumulatively explained 79% of variation in the rapeseed-mustard yield in Uttar Dinajpur.

Table 4.12 Summary of Stepwise Regression (SR) analysis for Rapeseed-mustard crop

District	Number of Steps	Predictor	Coefficient	p value
Cooch Behar	3	(Intercept)	-828.76	0.01
		CC_RH	2.38	0.05
		PC_Tmax	37.10	0.00
		CC_WS	220.67	0.02
Jalpaiguri	5	(Intercept)	617.09	0.06
		Year	12.81	0.00
		CC_Tmax	28.93	0.00
		PC_WS	-44.21	0.01
		Tmin	-41.03	0.04
		CC_WS	-129.54	0.05
Malda	4	(Intercept)	619.59	0.00
		Year	20.33	0.00
		CC_RF	2.43	0.00
		CC_RH	0.17	0.00
		PC_RH	3.13	0.03
Uttar Dinajpur	3	(Intercept)	-324.14	0.34
		Year	30.65	0.00
		CC_RH	12.76	0.00
		CC_WS	-128.66	0.01

From the results of stepwise regression, it can be observed that only two to seven explanatory variables are selected. The information present in the other variables is completely ignored. In order to achieve variable selection as well as to store maximum information present in the entire data in few of the selected variable simultaneously, two dimension reduction techniques namely, Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR) are employed.

4.3.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is conducted for each crop in all four districts by using their respective weather indices as well as time trend variable corresponding to the particular crop of a district. The variables are standardized before performing PCA. Sixteen Principal Components (PCs) are derived for each crop in each district as there are sixteen variables.

4.3.2.1 Rice

The summary of PCA for the rice crop of four districts is given in Table 4.13**.** The first principal component (PC_1) explains the largest variability present in the original data for all the districts. The amount of variability explained by the PC_1 is 28.18%, 37.29%, 32.03% and 34.45 for Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur respectively. The eigen value of the first five principal components are greater than one which indicates that each of the first five components explains more than one variable worth of information. Since each of those components explains the information that existed in more than one variable, the selection of such a component is beneficial in the view of dimension reduction. The first five principal components of Cooch Behar explain about 80% of the variation present in the original dataset. The first five components of Jalpaiguri, Malda and Uttar Dinajpur explained 80.90%, 84.44% and 83.91% of the variability original data respectively. It can be observed from the scree plot that is given in Fig. 4.5. The components after the fifth component explained less than 6% variability in the data. Hence, first five components of each district are selected for further analysis.

Table 4.13 Summary of Principal Component Analysis (PCA) for Rice crop

Parameters	**PC_1**	**PC_2**	**PC_3**	**PC_4**	**PC_5**
Cooch Behar					
Eigen value	4.51	2.61	2.22	2.03	1.32
Variability Explained (%)	28.18	16.29	13.87	12.67	8.24
Cumulative Variability (%)	28.18	44.46	58.34	71.00	79.24
Jalpaiguri					
Eigen value	5.97	2.20	1.96	1.64	1.18
Variability Explained (%)	37.29	13.73	12.26	10.25	7.37
Cumulative Variability (%)	37.29	51.02	63.28	73.53	80.90
Malda					
Eigen value	5.13	3.47	2.29	1.39	1.23
Variability Explained (%)	32.03	21.66	14.33	8.71	7.71
Cumulative Variability (%)	32.03	53.70	68.02	76.73	84.44
Uttar Dinajpur					
Eigen value	5.51	3.01	2.23	1.68	1.00
Variability Explained (%)	34.45	18.81	13.91	10.48	6.25
Cumulative Variability (%)	34.45	53.26	67.17	77.66	83.91

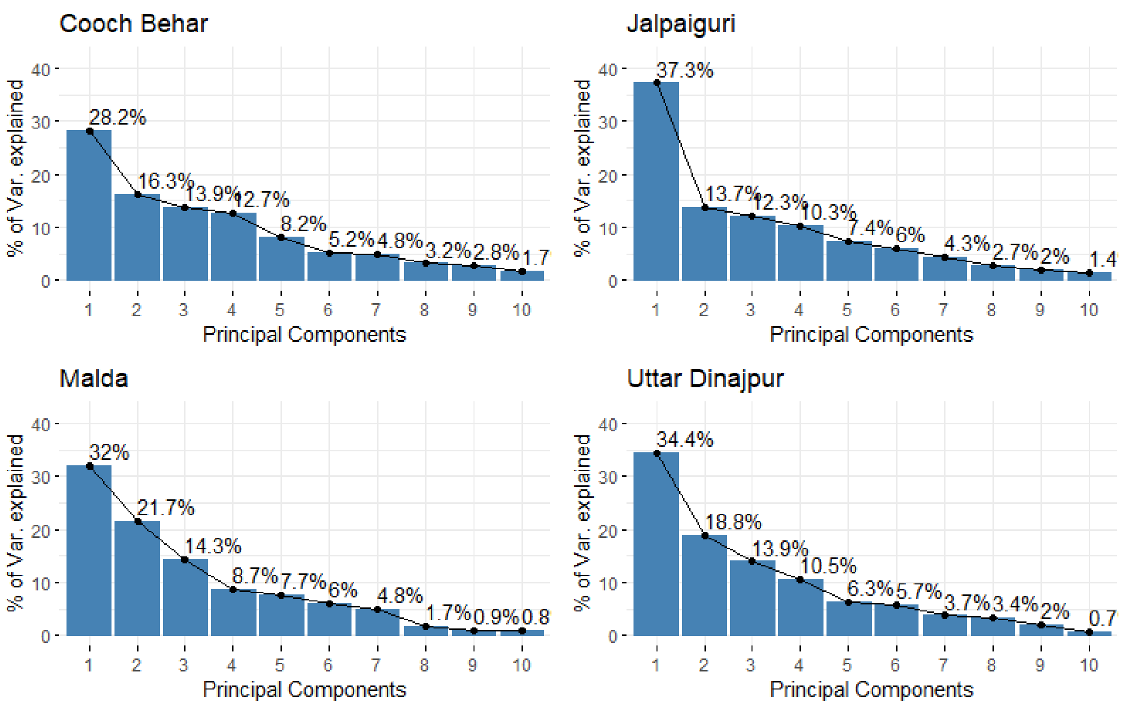

Fig. 4.6 Scree plot of Principal Component Analysis (PCA) for Rice crop

4.3.2.2 Wheat

The summary of PCA for wheat crop in four districts is given in Table 4.14. For Cooch Behar district, the first principal component (PC_1) is explaining the largest variability present in the original data (30.47%). The eigenvalues of the five components are greater than one. The first five components are cumulatively explained 82.60% of variability in the original data. From the scree plot given in Fig. 4.7, it can be observed that variability explained by the components after the fifth PCs is substantially low. Hence first five principal components that explained 82.60% variability of fifteen weather indices as well as time trend variable are selected.

Similarly, the eigenvalues of the first five PCs are greater than one in Malda and Uttar Dinajpur districts. The first five PCs are cumulatively explained 84.33% and 82.77% of variability in the original data respectively. It can be observed from the scree plot that the variability explained by the components after the fifth PCs are of lesser than 6%. For the Jalpaiguri district, the eigenvalues of first six PCs are greater than one and those explained about 87% of variability in the original data. Hence first five PCs of Cooch Behar, Malda and Uttar Dinajpur districts and first six PCs of Jalpaiguri are selected for further analysis.

Table 4.14 Summary of Principal Component Analysis (PCA) for Wheat crop

Parameters	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6
Cooch Behar						
Eigen value	4.87	3.27	2.21	1.48	1.38	0.83
Variability Explained (%)	30.47	20.43	13.81	9.24	8.65	5.18
Cumulative Variability (%)	30.47	50.90	64.71	73.95	82.60	87.77
Jalpaiguri						
Eigen value	5.04	2.99	2.32	1.39	1.16	1.01
Variability Explained (%)	31.52	18.68	14.47	8.70	7.23	6.30
Cumulative Variability (%)	31.52	50.20	64.67	73.38	80.61	86.91
Malda						
Eigen value	4.71	3.27	2.89	1.56	1.06	0.84
Variability Explained (%)	29.41	20.45	18.04	9.78	6.65	5.27
Cumulative Variability (%)	29.41	49.86	67.90	77.68	84.33	89.60
Uttar Dinajpur						
Eigen value	4.43	3.11	2.81	1.60	1.29	0.94
Variability Explained (%)	27.72	19.43	17.54	10.00	8.09	5.90
Cumulative Variability (%)	27.72	47.15	64.68	74.68	82.77	88.67

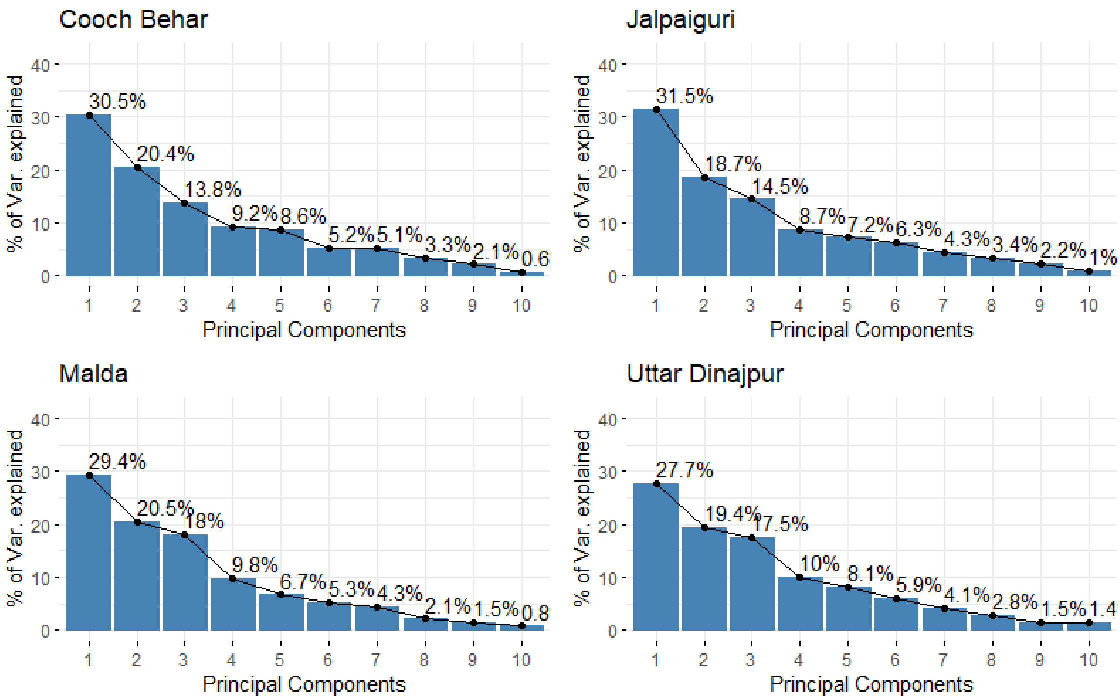

Fig. 4.7 Scree plot of Principal Component Analysis (PCA) for Wheat crop

4.3.2.3 Potato

The summary of PCA for potato crop in four districts is given in Table 4.15. The first principal component (PC_1) explains the largest variability present in the original data of all the districts. The amount of variability explained by the PC_1 is 29.69%, 29.00%, 35.10% and 39.43% for Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur respectively. The eigen values of first five principal components are greater than one for Cooch Behar, Jalpaiguri and Uttar Dinajpur districts which are cumulatively explaining 83.59%, 83.09% and 86.10% variability in the original data respectively. It can be observed from the scree plot that is given in Fig. 4.8. The components after fifth component explained less than 6% variability in the data.

Table 4.15 Summary of Principal Component Analysis (PCA) for Potato crop

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

The eigenvalues of the first four principal components are greater than one in Malda which are cumulatively explained 84.56% variability. The components after the fourth component explained substantially less variability. Hence first five components of Cooch Behar, Jalpaiguri and Uttar Dinajpur district and first four components of Malda are selected for further analysis.

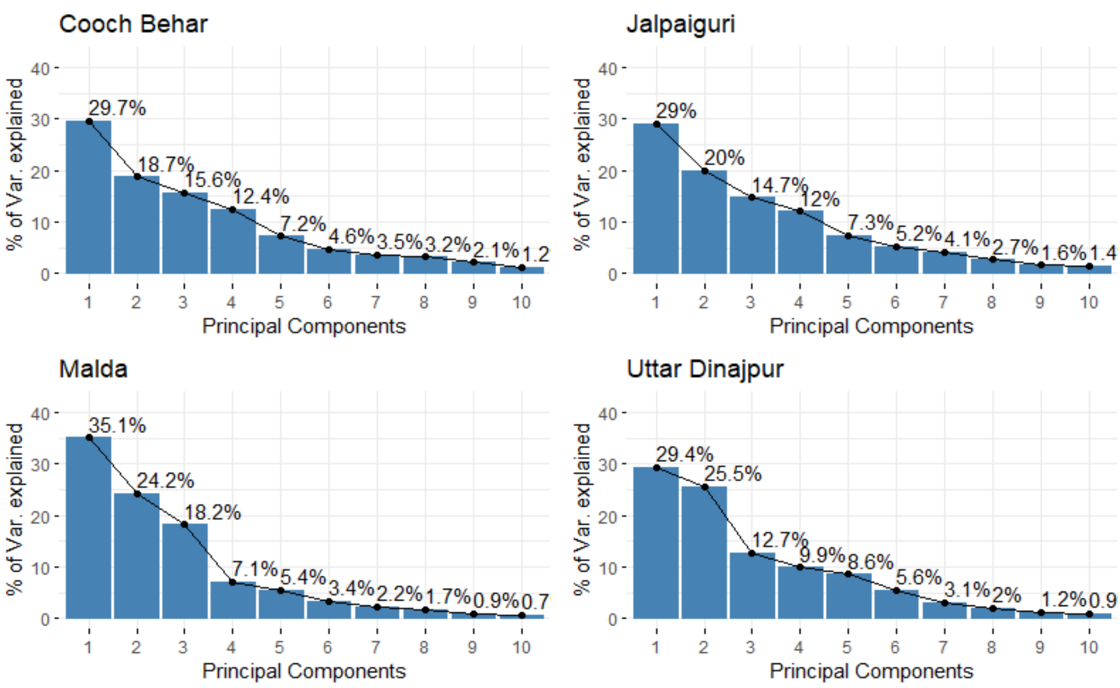

Fig. 4.8 Scree plot of Principal Component Analysis (PCA) for Potato crop

4.3.2.4 Jute

The summary of PCA for jute crop in four districts is given in Table 4.16. The amount of variability explained by the PC_1 is 39.60%, 26.25%, 27.30% and 37.14% in Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur respectively. The eigenvalues of the first four principal components are greater than one for Cooch Behar and Uttar Dinajpur districts while the first five components are having eigenvalues greater than one in Jalpaiguri and Malda districts. The first four components of Cooch Behar and Uttar Dinajpur are explaining 84.28% and 80.02% variability in the original data respectively. Similarly, the first five components of the Jalpaiguri and Malda districts explained 84.48% and 81.24% of the variability in the original dataset. It can be observed from the scree plot that is given in Fig. 4.9, the components after the fourth component of Cooch Behar and Uttar Dinajpur districts and components after the fifth component of Jalpaiguri and Malda are explaining substantially low variability in the data. Hence, the first four components of Cooch Behar and Uttar Dinajpur and the first five components of Jalpaiguri and Malda are selected for further modeling.

Table 4.16 Summary of Principal Component Analysis (PCA) for Jute crop

Parameters	**PC_1**	**PC_2**	**PC_3**	**PC_4**	**PC_5**
Cooch Behar					
Eigen value	6.34	3.28	2.46	1.41	0.93
Variability Explained (%)	39.60	20.52	15.35	8.81	5.83
Cumulative Variability (%)	39.60	60.12	75.47	84.28	90.11
Jalpaiguri					
Eigen value	4.20	3.71	2.40	2.14	1.07
Variability Explained (%)	26.25	23.19	14.99	13.35	6.70
Cumulative Variability (%)	26.25	49.44	64.43	77.79	84.48
Malda					
Eigen value	4.37	3.27	2.77	1.46	1.13
Variability Explained (%)	27.30	20.47	17.33	9.10	7.05
Cumulative Variability (%)	27.30	47.77	65.10	74.19	81.24
Uttar Dinajpur					
Eigen value	5.94	3.16	2.13	1.57	0.89
Variability Explained (%)	37.14	19.77	13.32	9.79	6.23
Cumulative Variability (%)	37.14	56.92	70.23	80.02	86.25

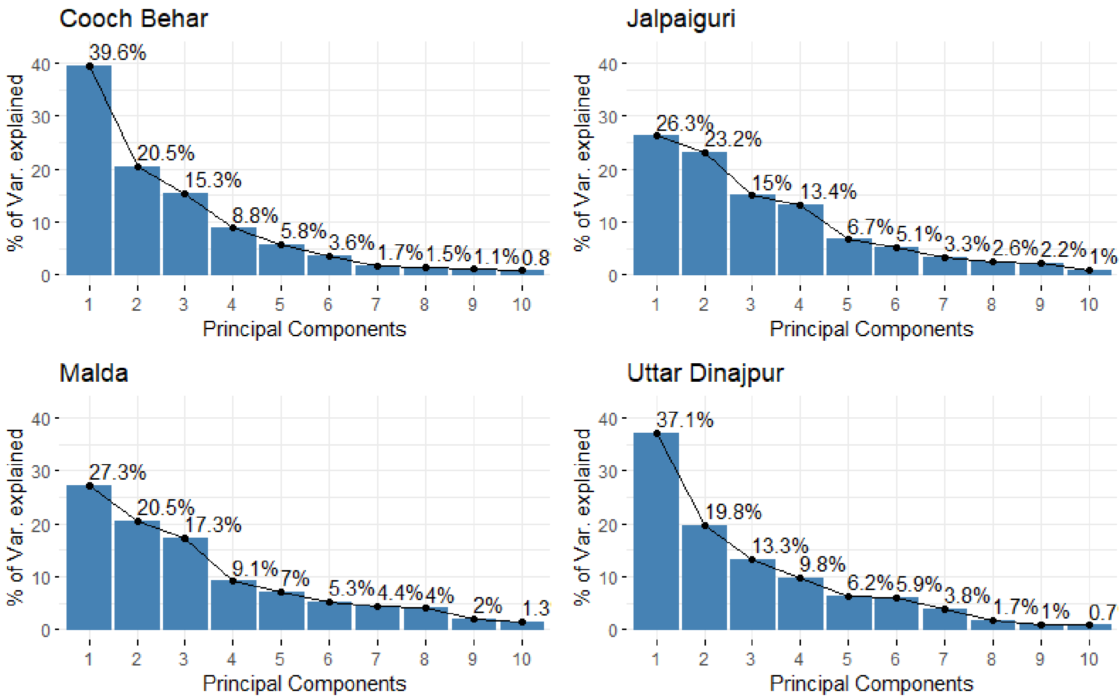

Fig. 4.9 Scree plot of Principal Component Analysis (PCA) for Jute crop

4.3.2.5 Rapeseed-mustard

The summary of PCA for rapeseed-mustard crop in four districts is given in Table 4.17. The first principal component (PC_1) explains the largest variability present in the original data for all the districts. The amount of variability explained by the PC_1 is 35.83%, 27.66%, 39.08% and 23.56% for Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur respectively. The eigenvalues of first five principal components are greater than one for all the districts. The first five principal components are cumulatively explaining 84.99%, 83.33%, 86.65% and 77.51% of the variability of the original data respectively

Table 4.17 Summary of Principal Component Analysis (PCA) for Rapeseed-mustard

Parameters	PC_1	PC_2	PC_3	PC_4	PC_5
Cooch Behar					
Eigen value	5.73	3.21	2.09	1.43	1.14
Variability Explained (%)	35.83	20.06	13.05	8.94	7.11
Cumulative Variability (%)	35.83	55.90	68.94	77.88	84.99
Jalpaiguri					
Eigen value	4.43	3.49	2.63	1.60	1.20
Variability Explained (%)	27.66	21.79	16.43	9.98	7.48
Cumulative Variability (%)	27.66	49.45	65.88	75.86	83.33
Malda					
Eigen value	6.25	3.13	1.83	1.59	1.05
Variability Explained (%)	39.08	19.59	11.44	9.96	6.59
Cumulative Variability (%)	39.08	58.66	70.11	80.06	86.65
Uttar Dinajpur					
Eigen value	3.77	3.24	2.33	1.69	1.37
Variability Explained (%)	23.56	20.26	14.55	10.58	8.57
Cumulative Variability (%)	23.56	43.82	58.37	68.95	77.51

. It can be observed from the scree plot that is given in Fig. 4.10, the components after fifth component are explaining substantially less variability of the original data.

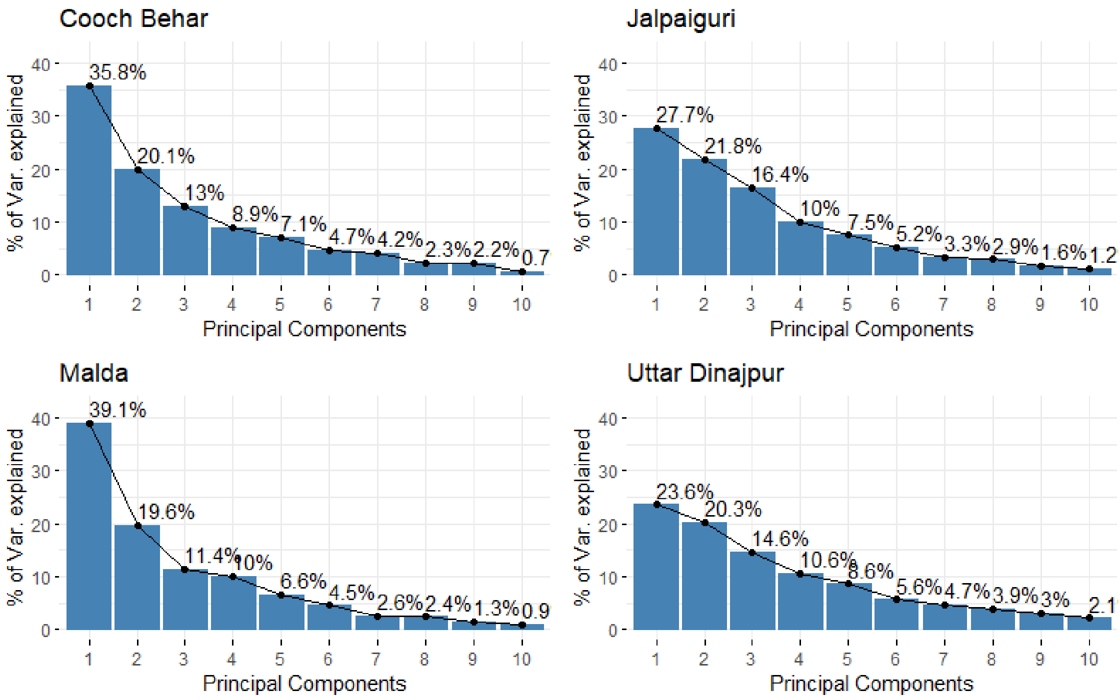

Fig. 4.10 Scree plot of Principal Component Analysis (PCA) for Rapeseed-mustard crop

4.3.3 Partial Least Square Regression (PLSR)

Partial Least Square Regression (PLSR) has been carried out for each crop of all four districts by taking crop yield as the dependent variable and their respective weather indices as well as time trend variable as independent variables. Sixteen partial least square components are obtained for each crop in each district as there are sixteen independent variables. The optimum number of components are selected using cross-validation technique. First few components up to a component for which the Root Mean Square Error for Cross Validation (RMSE-CV) is low are selected.

4.3.3.1 Rice

Cross-validation plots of PLSR for rice crop of each district are given in Fig. 4.11. For Cooch Behar district, the low RMSE-CV of 144.3 is obtained at the sixth component. The selection of the next component may increase the predictive error. Hence first six components are optimum. Similarly, the RMSE-CV is at the low level of 150.9 and 190.7 at the fifth components of the Jalpaiguri and Malda districts respectively. For Uttar Dinajpur, the low RMSE-CV of 167.6 is attained at the sixth component. Hence, the first six components of Cooch Behar and Uttar Dinajpur districts and the first five components of Jalpaiguri and Malda districts are selected for further analysis.

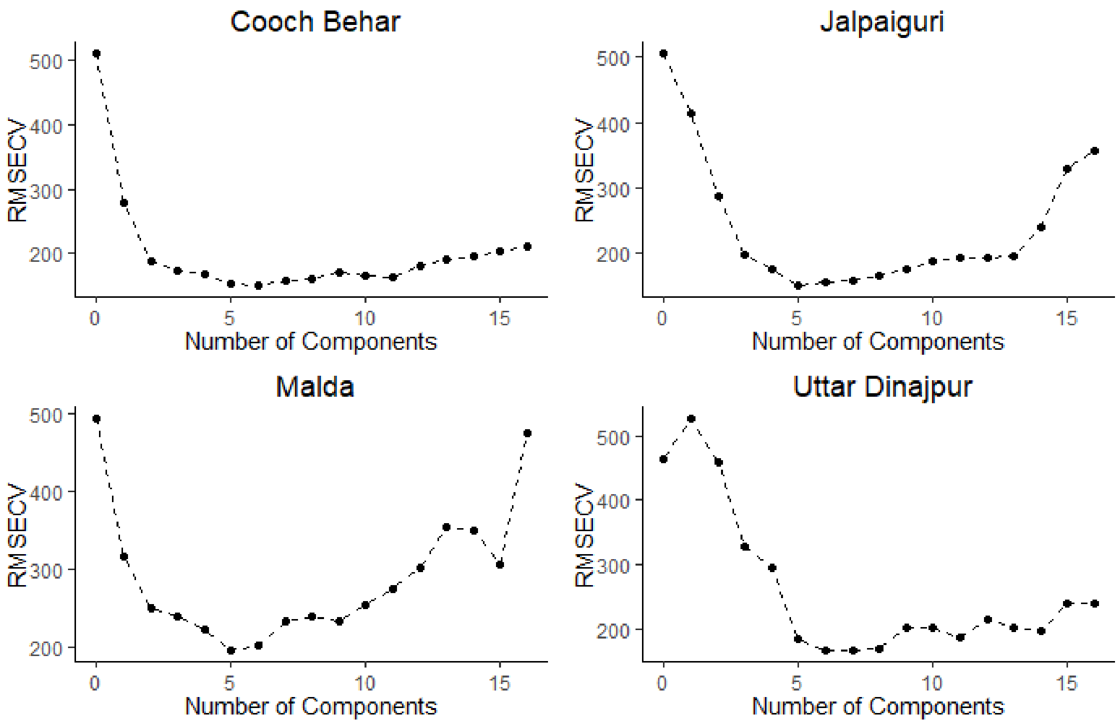

Fig. 4.11 PLSR cross-validation plots for Rice crop of four districts

The summary of PLSR for rice crop is given in Table 4.18. The first component of all the districts explained the highest variability in the yield. The first component of Cooch Behar is explaining 20.44% variability in the independent variables which can explain 82.21% of the variability in the yield. Similarly, the first component of Jalpaiguri, Malda and Uttar Dinajpur explained 28.63%, 24.17% and 12.82% variability in independent variables and they can explain 63.20%, 75.47% and 78.77% variability in yield respectively.

Table 4.18 Summary of Partial Least Square Regression (PLSR) for Rice crop

Cumulative Variability Explained	**Comp.1**	**Comp.2**	**Comp.3**	**Comp.4**	**Comp.5**	**Comp.6**
Cooch Behar						
Independent Variables (%)	20.44	42.22	51.49	61.18	64.22	73.12
Yield (%)	82.21	91.61	95.31	96.6	97.94	98.46
Jalpaiguri						
Independent Variables (%)	28.63	48.67	55.75	63.24	72.06	76.95
Yield (%)	63.20	89.85	96.20	97.99	98.74	99.16
Malda						
Independent Variables (%)	24.17	46.84	64.33	71.13	76.46	82.73
Yield (%)	75.47	86.49	89.46	93.91	95.15	95.41
Uttar Dinajpur						
Independent Variables (%)	12.82	40.23	55.03	67.76	77.98	85.64
Yield (%)	78.77	87.55	93.57	95.49	96.30	96.74

The first six components of the Cooch Behar and Uttar Dinajpur districts are explaining 73.12% and 85.64% variability in independent variables and they can explain 98.46% and 96.74% of the variability in the yield respectively. Similarly, the first five components of Jalpaiguri and Malda explained 72.06% and 76.46% of the variability in independent variables and they can explain 98.74% and 95.15% of the variability in the yield respectively. Hence, the PLSR components which explained the majority of variability in the independent variable also explained the maximum variability in the response variable (Duan *et al* 2020).

4.3.3.2 Wheat

Cross-validation plots of PLSR for wheat crop for each district are given in Fig. 4.12. For the Cooch Behar district, the RMSE-CV is low at a level of 304.0 at the fifth component. For Jalpaiguri and Uttar Dinajpur districts, the RMSE-CV is at the low level of 182.3 and 162.5 respectively at the fourth component. The low RMSE-CV of 176.6 is obtained at the seventh component for Malda. Hence, the first five components of Cooch Behar, seven components of Malda and four components of Jalpaiguri and Uttar Dinajpur districts are selected.

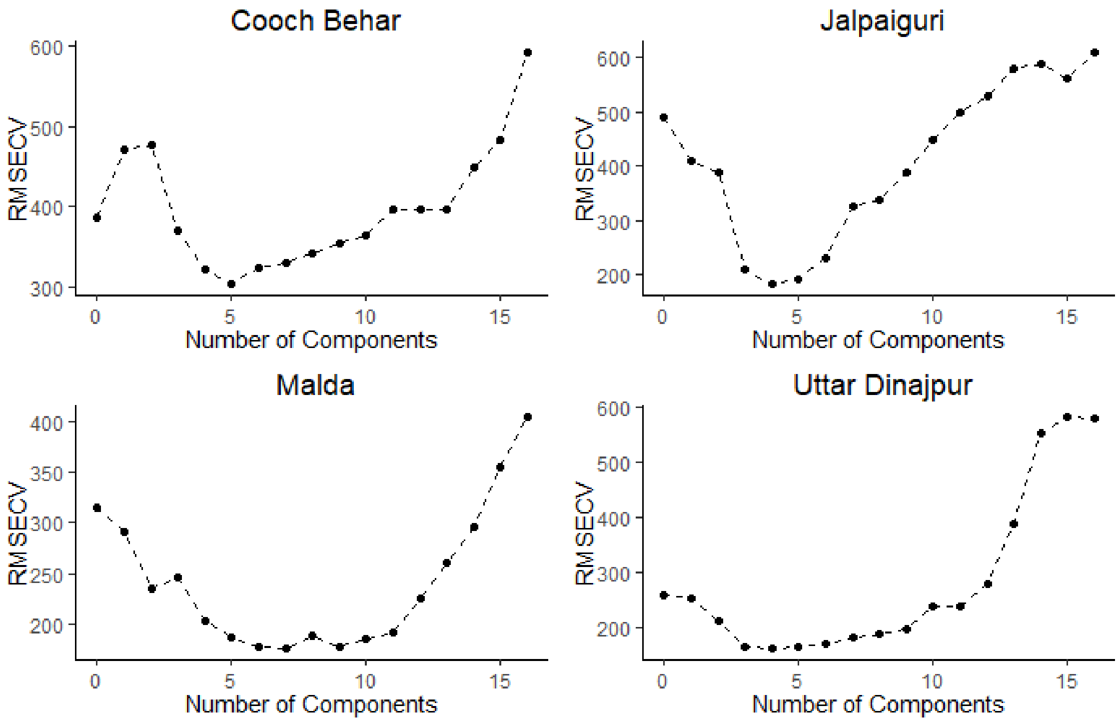

Fig. 4.12 PLSR cross-validation plots for Wheat crop of four districts

The summary of PLSR for wheat crop is given in Table 4.19. First component of Cooch Behar is explaining less variability in the independent variables (7.78%); but it can explain 69.73% of the variability in the yield. Similarly, first component of Jalpaiguri, Malda and Uttar Dinajpur explained 53.26%, 20.72% and 29.89% variability in independent variables and they are able to explain 97.51%, 90.40% and 38.88% variability in yield respectively. The first five components of Cooch Behar are cumulatively explaining 82.36% variability in independent variables and they are able to explain 99.77% of the variability in the yield. Similarly, first four components of Jalpaiguri and Uttar Dinajpur are cumulatively explaining 91.63% and 84.72% of variability in independent variables and they are able to explain 99.88% and 91.51% of variability in the yield respectively. First seven components that explain 93.05% of variability in independent variables are able to explain 99.74% of variability in the yield of wheat crop in Malda district.

Table 4.19 Summary of Partial Least Square Regression (PLSR) for Wheat crop

Cumulative Variability Explained	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Cooch Behar							
Independent Variables (%)	7.78	23.46	50.78	78.89	82.36	84.35	85.12
Yield (%)	69.73	84.67	94.48	99.57	99.77	99.83	99.93
Jalpaiguri							
Independent Variables (%)	53.26	61.46	90.48	91.63	92.92	93.23	94.07
Yield (%)	97.51	99.41	99.79	99.88	99.94	99.97	99.98
Malda							
Independent Variables (%)	20.72	65.98	76.28	80.90	89.17	89.51	93.05
Yield (%)	90.40	94.05	95.19	97.00	97.63	99.70	99.74
Uttar Dinajpur							
Independent Variables (%)	29.89	69.23	78.29	84.72	87.33	88.25	89.99
Yield (%)	38.88	55.57	90.03	91.51	93.79	96.65	97.61

4.3.3.2 Potato

Cross-validation plots of PLSR for potato crop for each district are given in the Fig. 4.13. The low RMSE-CV of 5.19 is obtained at first component of Cooch Behar. The inclusion of the second component increases the predictive error. Hence only the first component of Cooch Behar district is selected. For Jalpaiguri and Uttar Dinajpur districts, the low RMSE-CV of 2.82 and 3.13 are obtained at the fourth component respectively. For Malda, the RMSE-CV is at a low level of 6.10 at the fifth component. Hence, the first four components of Jalpaiguri and Uttar Dinajpur districts and five components of Malda are selected.

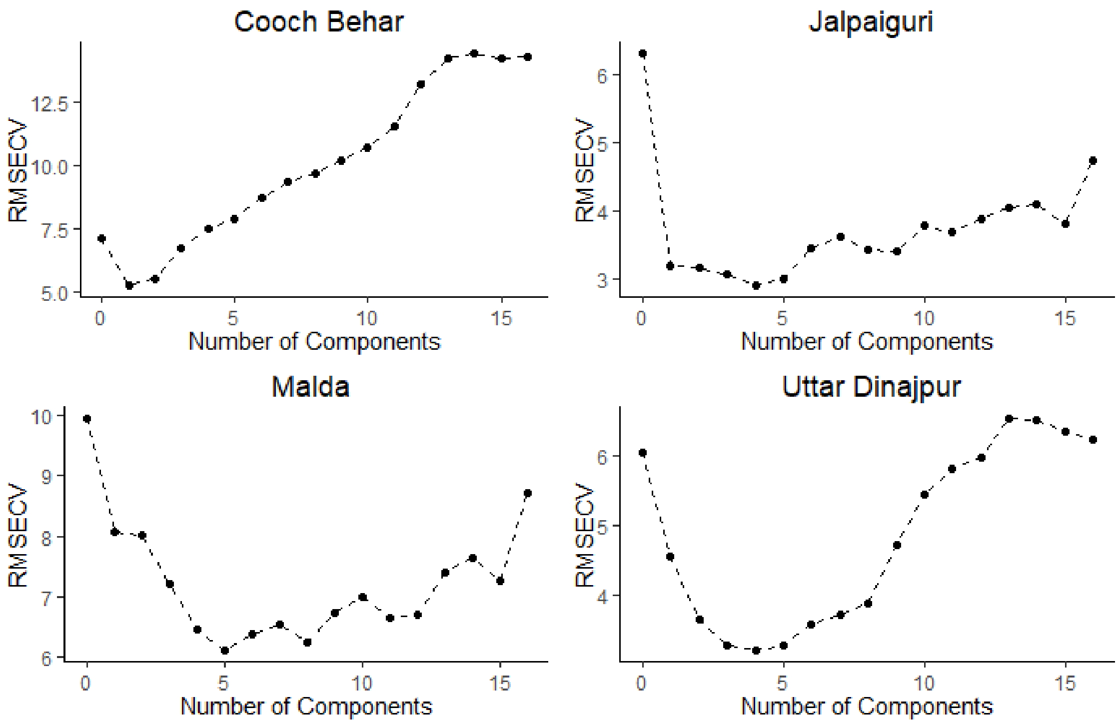

Fig. 4.13 PLSR cross-validation plots for Potato crop of four districts

The summary of PLSR for potato crop is given in Table 4.20. The first component of Cooch Behar is explaining 24.68% variability in the independent variables which is able to explain 66.48% of the variability in the yield. Similarly, the first component of Jalpaiguri, Malda and Uttar Dinajpur explained 28.50%, 34.54% and 27.12% variability in independent variables and they are able to explain 81.16%, 61.47% and 72.52% variability in yield respectively.

Table 4.20 Summary of Partial Least Square Regression (PLSR) for Potato crop

Cumulative Variability Explained	**Comp.1**	**Comp.2**	**Comp.3**	**Comp.4**	**Comp.5**
Cooch Behar					
Independent Variables (%)	24.68	46.5	56.06	63.8	67.72
Yield (%)	66.48	73.1	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

First four components of Jalpaiguri and Uttar Dinajpur are cumulatively explaining 58.46% and 72.43% of variability in independent variables and they are able to explain 94.30% and 90.86% of variability in the yield respectively. Similarly, first five components of Malda are cumulatively explained 86.39% variability in independent variables and they are able to explain 89.12% of variability in the yield.

4.3.3.4 Jute

PLSR cross-validation plots for jute crop for each district are given in Fig. 4.14. The low RMSE-CV of 0.70 and 1.22 are obtained at the fourth component of Cooch Behar and Malda districts respectively. Similarly, RMSE-CV is at a low level of 0.85 and 0.65 at the sixth and fifth components of the Jalpaiguri and Uttar Dinajpur districts respectively. Hence, the first four components of Cooch Behar and Malda districts, six components of Jalpaiguri and five components of Uttar Dinajpur are selected for further analysis.

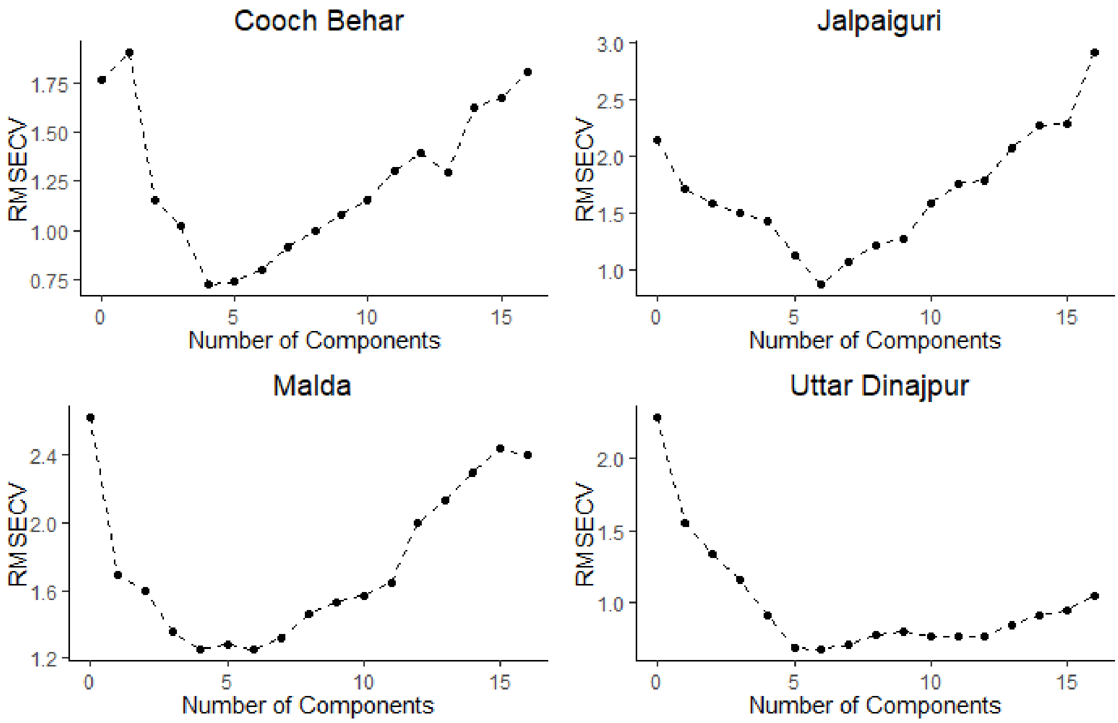

Fig. 4.14 PLSR cross-validation plots for Jute crop of four districts

The summary of PLSR for jute crop is given in Table 4.21. The first component of Cooch Behar is explaining less variability in the independent variables (17.29%); but it can explain 70.29% of variability in the yield. Similarly, the first component of Jalpaiguri, Malda and Uttar Dinajpur explained 21.16%, 20.69% and 27.02% variability in independent variables and they can explain 69.30%, 74.22% and 70.67% variability in yield respectively. First four components of Cooch Behar and Malda are cumulatively explaining 80.03% and 67.81% of the variability in independent variables and they are able to explain 92.59% and 92.19% of variability in the yield respectively. Similarly, first six components of Jalpaiguri are cumulatively explaining 85.26% of variability in independent variables and they are able to explain 95.44% of variability in the yield. First five components that explain 76.32% of variability in independent variables are able to explain 93.74% of variability in the yield of jute crop in Uttar Dinajpur district.

Table 4.21 Summary of Partial Least Square Regression (PLSR) for Jute crop

Cumulative Variability Explained	**Comp.1**	**Comp.2**	**Comp.3**	**Comp.4**	**Comp.5**	**Comp.6**
Cooch Behar						
Independent Variables (%)	17.29	44.62	64.59	80.03	85.89	91.68
Yield (%)	70.29	82.38	90.36	92.59	94.83	95.75
Jalpaiguri						
Independent Variables (%)	21.16	36.04	47.20	65.66	73.16	85.26
Yield (%)	69.30	82.84	91.69	94.32	95.17	95.44
Malda						
Independent Variables (%)	20.69	38.18	52.12	67.81	73.95	80.77
Yield (%)	74.22	84.89	91.15	92.19	93.74	94.30
Uttar Dinajpur						
Independent Variables (%)	27.02	53.75	64.48	73.69	76.32	81.76
Yield (%)	70.67	84.16	93.26	97.00	97.95	98.05

4.3.3.5 Rapeseed-mustard

PLSR cross-validation plots for the rapeseed-mustard crop for each district are given in Fig. 4.15. The low RMSE-CV of 96.66 is obtained at second component of Cooch Behar. Similarly, the RMSE-CV is at low level of 69.30 and 84.67 at fourth component of Jalpaiguri and Malda districts respectively. For Uttar Dinajpur, the low RMSE-CV of 151.50 is attained at first component. Hence, the first two components of Cooch Behar, four components of Jalpaiguri and Malda districts and only first component of Uttar Dinajpur district are selected for further analysis.

The summary of PLSR for the rapeseed-mustard crop is given in Table 4.22. First component of Cooch Behar is explaining 34.77% variability in the independent variables which is able to explain 67.11% of variability in the yield. Similarly, first components of Jalpaiguri and Malda are explaining 23.59% and 23.51% variability in independent variables and they are able to explain 80.28% and 81.08% variability in yield respectively. First two components of Cooch Behar are cumulatively explaining 51.63% variability in independent variables which are able to explain 74.18% of variability in the yield. First four components of Jalpaiguri and Malda are cumulatively explaining 62.25% and

72.87% of variability in independent variables and they are able to explain 91.42% and 94.06% of variability in the yield respectively. First component of Uttar Dinajpur district is explaining 23.17% variability in independent variables and they are able to explain 72.21% of variability in the yield.

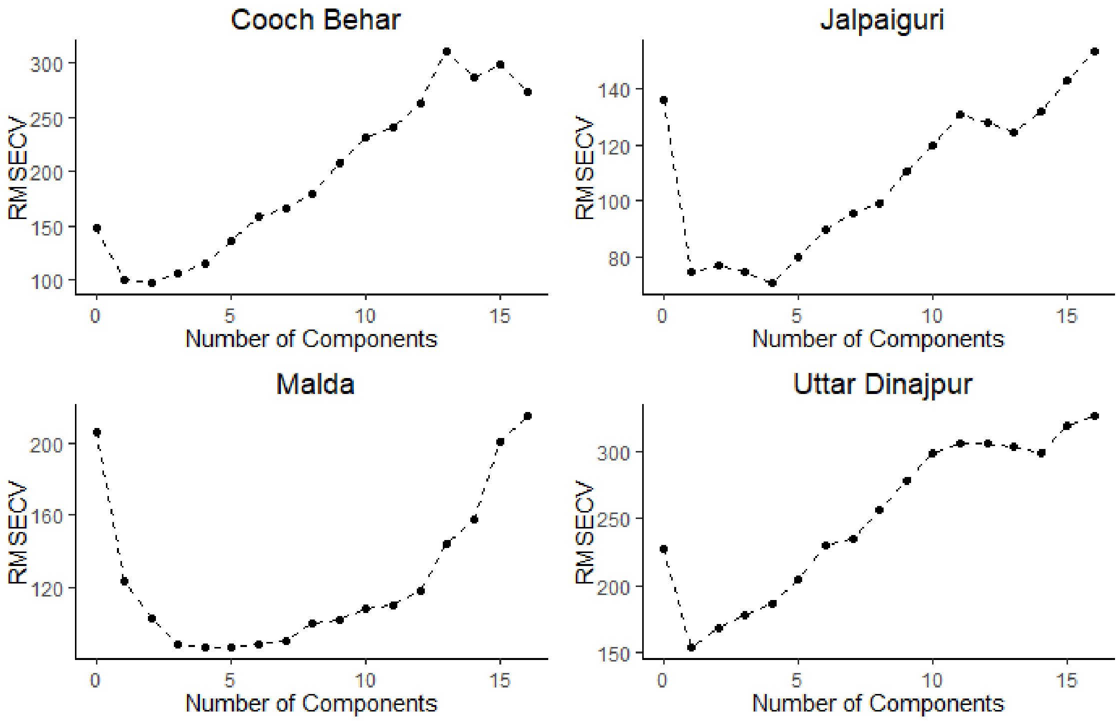

Fig. 4.15 PLSR cross-validation plots for Rapeseed-mustard crop of four districts

Table 4.22 Summary of Partial Least Square Regression (PLSR) for Rapeseed-Mustard

Cumulative Variability Explained	**Comp.1**	**Comp.2**	**Comp.3**	**Comp.4**
Cooch Behar				
Independent Variables (%)	34.77	51.63	63.29	72.89
Yield (%)	67.11	74.18	77.5	79.69
Jalpaiguri				
Independent Variables (%)	23.59	43.57	56.12	62.25
Yield (%)	80.28	85.81	89.62	91.42
Malda				
Independent Variables (%)	23.51	56.99	67.1	72.87
Yield (%)	81.08	87.65	91.23	94.06
Uttar Dinajpur				
Independent Variables (%)	23.17	32.47	46.01	56.96
Yield (%)	72.71	79.1	80.88	82.33

4.4 Fitting and Estimation of parameters of different Statistical and Machine Learning Models

Multiple Linear Regression (MLR), three penalized regression models *viz.* Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) and two machine learning models *viz.* Artificial Neural Network (ANN) and Support Vector Regression (SVR) have been fitted for each crop in each district. The MLR, ANN and SVR model have been fitted separately under three variable selection or dimension reduction methodologies *viz.* SR, PCA and PLSR. While fitting the models, the crop yield is taken as the dependent variable or output variable and the variables selected from SR or components selected from PCA or PLSR as the explanatory variables or input variables.

The penalized regression models are doing inner variable selection by shrinking the coefficient values. As a result of inner variables selection, only important variables are retained in the model and due to coefficient shrinkage, the overfitting issue is also mitigated. Hence, there is no necessity of variable selection or dimension reduction in penalized regression models. Therefore, while fitting the penalized regression models, the crop yield is taken as the dependent variable and all the fifteen weather indices as well as time trend variable as independent variables or explanatory variables.

4.4.1 Multiple Linear Regression (MLR) Model

4.4.1.1 Stepwise Regression-Multiple linear Regression (SR-MLR)

In SR-MLR, the variables selected from the stepwise regression are used as the independent variables and the yield of the respective crop in the respective district are used as the dependent variable. Since the parameters of SR-MLR are estimated while employing the stepwise regression in section 4.3.1, here it is not included again in order to avoid ambiguity.

4.4.1.2 Principal Component Analysis-Multiple linear Regression (PCA-MLR)

In PCA-MLR, the selected principal components of a respective crop in a respective district are used as the independent variables and crop yield is taken as the dependent variable.

4.4.1.2 a) Rice

Five principal components were optimum for all four districts of rice crop. Hence, the regression models have been fitted using five principal components as independent variables and the rice crop yield of the respective district as the dependent variable. The summary of PCA-MLR for rice crop is given in Table 4.23. It can be seen that the first two principal components (PCs) and fourth PC are highly significant at 1% level of significance (LoS) for the Cooch Behar district. For Jalpaiguri, PC_1, PC_3 and PC_5 are highly significant at 1% LoS and PC_2 is significant at 5% LoS. First three PCs of Malda are significantly influenced the yield of rice crop. For Uttar Dinajpur, only PC_4 is significant and other components are non-significant at 5% LoS.

Table 4.23 Summary of PCA-MLR model for Rice crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PCA Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	1963.99	0.00	1904.37	0.00	2620.17	0.00	2167.51	0.00
PC_1	96.81	0.00	81.25	0.01	-80.50	0.01	16.49	0.61
PC_2	199.78	0.00	-91.61	0.03	-107.76	0.01	34.31	0.53
PC_3	24.69	0.49	-234.95	0.00	218.14	0.00	85.79	0.15
PC_4	163.61	0.00	1.88	0.97	-23.41	0.70	-257.12	0.00
PC_5	-33.87	0.56	-181.90	0.00	-26.15	0.68	-12.09	0.89

4.4.1.2 b) Wheat

The first five principal components were optimum for Cooch Behar, Malda and Uttar Dinajpur districts and it is six for Jalpaiguri. The summary of PCA-MLR for wheat crop is given in Table 4.24. It can be seen that only first two principal components (PCs) are highly significant at 1% LoS for the Cooch Behar district and the remaining components are non-significant at 5% LoS. For Jalpaiguri, PC_1 is highly significant at 1% LoS and PC_2, PC_3 and PC_5 are significant at 5% LoS. PC_1, PC_4 and PC_5 are significant at 1% LoS for Malda. For Uttar Dinajpur, only PC_2 and PC_4 are significant and other components are non-significant at 5% LoS.

Table 4.24 Summary of PCA-MLR model for Wheat crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PCA Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	2041.56	0.00	2132.84	0.00	2752.15	0.00	2390.89	0.00
PC_1	98.66	0.01	171.91	0.00	61.92	0.01	-19.99	0.21
PC_2	-113.91	0.00	-92.74	0.02	21.61	0.36	-92.10	0.02
PC_3	29.41	0.46	-87.23	0.02	28.40	0.25	66.44	0.01
PC_4	-43.73	0.39	-1.71	0.97	-129.68	0.00	-75.98	0.06
PC_5	33.51	0.48	108.02	0.04	115.59	0.01	47.30	0.21
PC_6	NA	NA	-5.86	0.92	NA	NA	NA	NA

4.4.1.2 c) Potato

The first five principal components were optimum for Cooch Behar, Jalpaiguri and Uttar Dinajpur districts and it is four for Malda. The summary of PCA-MLR for potato crop is given in the Table 4.25. It can be seen that only the first two principal components (PCs) are highly significant at 1% LoS for Cooch Behar district and the remaining components are non-significant at 5% LoS. For Jalpaiguri, first four components are highly significant at 1% LoS. PC_1 is significant at 1% LoS and PC_5 is significant at 5% LoS for Malda. For Uttar Dinajpur, the first three components as well as PC_5 are significant at 1% LoS.

Table 4.25 Summary of PCA-MLR model for Potato crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PCA Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	24.30	0.00	41.28	0.00	26.79	0.00	23.51	0.00
PC_1	1.48	0.01	-2.65	0.00	2.66	0.00	1.46	0.00
PC_2	2.33	0.00	0.55	0.00	-0.66	0.40	-1.46	0.00
PC_3	-0.28	0.67	-0.14	0.00	1.28	0.16	1.32	0.01
PC_4	0.98	0.28	0.70	0.01	-2.94	0.05	-0.45	0.41
PC_5	0.00	1.00	1.71	0.35	NA	NA	-2.14	0.00

4.4.1.2 d) Jute

The first four principal components of Cooch Behar and Uttar Dinajpur and the first five components of Jalpaiguri and Malda are found to be optimum. The summary of PCA-MLR for the jute crop is given in Table 4.26.

Table 4.26 Summary of PCA-MLR model for Jute crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PCA Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	11.31	0.00	11.91	0.00	14.07	0.00	11.95	0.00
PC_1	-0.06	0.60	-0.29	0.09	0.19	0.26	0.35	0.01
PC_2	0.33	0.06	0.55	0.00	-0.73	0.00	-0.77	0.00
PC_3	-0.52	0.01	-0.25	0.23	0.60	0.01	-0.69	0.00
PC_4	0.46	0.08	-0.71	0.00	-0.04	0.90	-0.04	0.86
PC_5	NA	NA	-0.28	0.35	0.43	0.32	NA	NA

It can be seen that only PC_1 is highly significant at 1% LoS for Cooch Behar district and remaining components are non-significant at 5% LoS. For Jalpaiguri, PC_2 and PC_4 are significant at 1% LoS. PC_2 and PC_3 are significant at 1% LoS for Malda. For Uttar Dinajpur, first three components are significant at 1% LoS.

4.4.1.2 e) Rapeseed-mustard

First five principal components have been found to be optimum for all the four districts. The summary of PCA-MLR for rapeseed-mustard crop is given in the Table 4.27. For Cooch Behar district, only PC_1 and for Jalpaiguri, PC_1 and PC_2 are significantly influencing the rapeseed-mustard yield at 1% LoS. PC_2 of Malda is significant at 1% LoS as well as PC_4 and PC_5 are significant at 5% LoS. For Uttar Dinajpur, only first component is significant at 1% LoS.

Table 4.27 Summary of PCA-MLR model for Rapeseed-Mustard crop

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PCA Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	545.16	0.00	624.65	0.00	1062.70	0.00	820.78	0.00
PC_1	49.52	0.00	-26.97	0.00	14.15	0.08	95.86	0.00
PC_2	25.58	0.07	-54.69	0.00	-93.26	0.00	-33.42	0.09
PC_3	27.79	0.09	0.43	0.97	-7.40	0.64	-26.27	0.21
PC_4	14.57	0.45	20.27	0.16	-37.04	0.02	-12.86	0.60
PC_5	1.59	0.94	-24.16	0.12	-51.80	0.03	27.16	0.29

4.4.1.3 Partial Least Square Regression-Multiple linear Regression (PLSR-MLR)

In PLSR-MLR, the selected PLSR components of a respective crop in a respective district are used as the independent variables and crop yield is taken as the dependent variable.

4.4.1.3 a) Rice

The first six PLSR components of Cooch Behar and Uttar Dinajpur and first five components of Jalpaiguri and Malda are found to be optimum. The summary of PLSR-MLR for rice crop is given the Table 4.28. It can be observed that the first five PLSR components out of six components were significantly influencing the rice yield of Cooch Behar. Similarly, first four components out of five components of Jalpaiguri and Malda are significant at 1% LoS. The first three components out of six components are significant at 1% LoS for Uttar Dinajpur. Since the PLSR components are derived in such a way that it maximizes covariance between the independent variables set and dependent variable, the derived components are significantly influencing the yield.

Table 4.28 Summary of PLSR-MLR model for Rice crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PLSR Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	1956.80	0.00	1927.49	0.00	2620.91	0.00	2203.17	0.00
Comp.1	267.24	0.00	212.40	0.00	223.85	0.00	307.28	0.00
Comp.2	89.98	0.00	145.48	0.00	85.70	0.00	73.41	0.00
Comp.3	88.26	0.00	131.18	0.00	71.08	0.01	100.97	0.00
Comp.4	54.28	0.00	78.17	0.00	104.75	0.00	47.26	0.08
Comp.5	114.00	0.00	31.39	0.10	71.48	0.10	47.64	0.08
Comp.6	39.14	0.12	N/A	N/A	N/A	N/A	31.00	0.20

4.4.1.3 b) Wheat

The First five PLSR components of Cooch Behar, four components of Jalpaiguri and Uttar Dinajpur districts and first seven components of Malda were found to be optimum. The summary of PLSR-MLR for wheat crop is given in Table 4.29. It can be observed that the first PLSR component is significant at 5% LoS and second, third and fourth components are significant at 1% LoS for Cooch Behar. Similarly, first three components out of four components of Jalpaiguri and first five components out of seven components of Malda are significant at 1% LoS. For Uttar Dinajpur, first two components and fourth component are significant at 1% LoS and third component is significant at 5% LoS.

Table 4.29 Summary of PLSR-MLR model for Wheat crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PLSR Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	2052.03	0.00	2151.09	0.00	2750.15	0.00	2393.02	0.00
Comp.1	1.16	0.05	1.39	0.00	2.19	0.00	9.93	0.00
Comp.2	6.29	0.00	7.36	0.00	14.72	0.00	19.12	0.00
Comp.3	8.91	0.00	16.13	0.00	12.79	0.00	4.35	0.02
Comp.4	8.71	0.00	14.86	0.12	11.16	0.01	31.29	0.01
Comp.5	9.21	0.39	N/A	N/A	14.74	0.01	N/A	N/A
Comp.6	N/A	N/A	N/A	N/A	0.94	0.71	N/A	N/A
Comp.7	N/A	N/A	N/A	N/A	41.29	0.02	N/A	N/A

4.4.1.3 c) Potato

Only the first PLSR component of Cooch Behar, first four components of Jalpaiguri and Uttar Dinajpur districts and the first five components of Malda were found to be optimum. The summary of PLSR-MLR for potato crop is given in Table 4.30. It can be observed that the first PLSR component of Cooch Behar is highly significant at 1% LoS. First three components of Jalpaiguri and Malda are significant at 1% LoS. Similarly, two components out of four components of Uttar Dinajpur are significant at 1% LoS.

Table 4.30 Summary of PLSR-MLR model for Potato crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PLSR Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	24.39	0.00	24.98	0.00	27.12	0.00	23.63	0.00
Comp.1	2.75	0.00	2.58	0.00	3.18	0.00	2.52	0.00
Comp.2	N/A	N/A	1.57	0.00	3.88	0.00	1.78	0.00
Comp.3	N/A	N/A	1.26	0.01	1.92	0.01	0.41	0.27
Comp.4	N/A	N/A	0.55	0.25	0.88	0.31	0.73	0.10
Comp.5	N/A	N/A	N/A	N/A	1.34	0.10	N/A	N/A

4.4.1.3 d) Jute

The first four PLSR components of Cooch Behar and Malda districts, six components of Jalpaiguri and the first five components of Uttar Dinajpur were found to be optimum. The summary of PLSR-MLR for jute crop is given in Table 4.31.

Table 4.31 Summary of PLSR-MLR model for Jute crop in four districts

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PLSR Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	11.25	0.00	11.79	0.00	14.00	0.00	11.78	0.00
Comp.1	0.92	0.00	0.98	0.00	1.20	0.00	0.95	0.00
Comp.2	0.35	0.00	0.57	0.00	0.63	0.00	0.44	0.00
Comp.3	0.28	0.00	0.62	0.00	0.42	0.01	0.65	0.00
Comp.4	0.19	0.05	0.21	0.02	0.16	0.30	0.39	0.00
Comp.5	N/A	N/A	0.20	0.15	N/A	N/A	0.32	0.03
Comp.6	N/A	N/A	0.08	0.48	N/A	N/A	N/A	N/A

It can be observed that the first three PLSR components of Cooch Behar and Jalpaiguri are significantly influencing the jute yield of respective districts at 1% LoS and the fourth component is also significant at 5% LoS. Similarly, the first three components out of four components of Malda are significant at 1% LoS. The first four components are significant at 1% LoS and fifth component also significant at 5% LoS for Uttar Dinajpur.

4.4.1.3 e) Rapeseed-mustard

Only two PLSR components of Cooch Behar, the first four components of Jalpaiguri and Malda districts are found to be optimum. In the case of Uttar Dinajpur, only the first component is found to be optimum. The summary of PLSR-MLR for rapeseed-mustard crop is given in Table 4.32. For Cooch Behar, the first component is highly significant at 1% LoS and the second component is also significant at 5% LoS. Similarly, first two components of Jalpaiguri are significant at 1% LoS and the third component is also significant at 5% LoS. All four components of Malda are significantly influencing the rapeseed-mustard yield at 1% LoS. Only one PLSR component which is found to be sufficient is significantly influencing the rapeseed-mustard yield of Uttar Dinajpur at 1% LoS.

Table 4.32 Summary of PLSR-MLR model for Rapeseed-Mustard crop

District	Cooch Behar		Jalpaiguri		Malda		Uttar Dinajpur	
PLSR Components	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**	**Regression Coefficient**	**p value**
(Intercept)	548.31	0.00	624.74	0.00	1058.69	0.00	816.03	0.00
Comp.1	56.29	0.00	63.31	0.00	95.16	0.00	104.18	0.00
Comp.2	34.02	0.02	24.33	0.01	24.77	0.00	N/A	N/A
Comp.3	N/A	N/A	18.84	0.04	34.37	0.01	N/A	N/A
Comp.4	N/A	N/A	20.60	0.09	53.25	0.01	N/A	N/A

It can be observed that most of the PLSR components of PLSR-MLR models are significant. The earlier PLSR components are significantly influencing the response variable since the PLSR components are derived by considering the response variable also. Only a few last components are non-significant. But in PCA based model there is no such order. Theoretically, the first few principal components may explain almost close to 100% variability in the explanatory variable. But there is no surety that the components that explain the majority of the variability in the explanatory variable, explain the greater variability in the response variable. The last principal component that explains less variability in the explanatory variables set may fit the response very well; but it is generally ignored (Hadi and Ling, 1998).

4.4.2 Artificial Neural Network (ANN) Model

Three ANN models have been fitted for each crop in each district by taking indices selected from stepwise regression as input (SR-ANN), principal components as input (PCA-ANN) and PLSR components as input (PLSR-ANN). The Multilayer Perceptron (MLP) architecture of the ANN model is trained using the Backpropagation (BP) algorithm. The performance of three activation functions *viz.* Logistic, Tangent Hyperbolic (Tanh) and Restricted Linear Unit (ReLU) have been evaluated.

The number of input layer neurons of SR-ANN, PCA-ANN and PLSR-ANN models are the number of indices selected using SR, the number of principal components selected from PCA and the number of components selected from PLSR respectively. The number of output layer neurons is one as there is only one output variable which is crop yield. The number of neurons in the hidden layer is determined using 10-fold cross-validation technique. The number of neurons in the hidden layer for which the Root Mean Square Error for Cross Validation (RMSE-CV) is low are considered as the optimum choice. The data is normalized before training ANN model in order to avoid bias due to the magnitude of the input variable and its range of variation.

4.4.2.1 Rice

The results of SR-ANN, PCA-ANN and PLSR-ANN models based on the backpropagation algorithm using three different activation functions for rice crop of all four districts are given in the Table 4.33.

4.4.2.1 a) Cooch Behar

The RMSE-CV is low for six hidden layer neurons for PCA-ANN and PLSR-ANN models and it is one for SR-ANN model. The number of input neurons for SR-ANN is six since six weather indices are found to have a significant effect on rice yield of Cooch Behar in stepwise regression. Similarly, the number of input neurons was five and six for PCA-ANN and PLSR-ANN as the first five and six components of PCA and PLSR respectively are found to be optimal.

In the SR-ANN model, the ReLU activation function is converged to the lowest possible Sum of Squared Error (SSE) of 0.01 with a learning rate of 0.03 in 591 iterations. The learning rate of Tanh is also 0.03 and it converged to SSE of 0.02 in 236 steps. However, the SSE is low in ReLU. Further, the AIC value is also low in ReLU. Hence the ReLU function is found to perform better in SR-ANN. In PCA-ANN, the logistic activation function is converged to the lowest possible SSE of 0.11 with a learning rate of 0.07. But it has taken many iterations to converge.

Table 4.33 Summary of ANN models using three activation functions for Rice crop in four districts

District	Model	Number of Input Neurons	Number of Hidden layer Neurons	Activation Function	Learning Rate	SSE	Steps	AIC
Cooch Behar	SR-ANN	6	6	Logistic	0.07	0.03	1120	98.05
				Tanh	0.03	0.02	236	98.04
				ReLU	0.03	0.01	591	98.03
	PCA-ANN	5	1	Logistic	0.07	0.11	18298	16.22
				Tanh	0.07	0.12	1610	16.23
				ReLU	0.09	0.79	39	17.59
	PLSR-ANN	6	6	Logistic	0.09	0.80	57	99.59
				Tanh	0.02	0.02	868	98.03
				ReLU	0.09	0.79	52	99.59
Jalpaiguri	SR-ANN	5	5	Logistic	0.06	0.02	888	72.03
				Tanh	0.03	0.01	795	72.02
				ReLU	0.09	0.96	47	73.92
	PCA-ANN	5	2	Logistic	0.07	0.17	3583	30.33
				Tanh	0.07	0.18	480	30.36
				ReLU	0.09	0.92	70	31.84
	PLSR-ANN	5	5	Logistic	0.09	0.94	47	73.89
				Tanh	0.03	0.02	360	72.03
				ReLU	0.09	0.96	45	73.92
Malda	SR-ANN	5	5	Logistic	0.06	0.07	671	72.14
				Tanh	0.02	0.06	974	72.13
				ReLU	0.09	0.76	47	73.53
	PCA-ANN	5	5	Logistic	0.09	0.77	47	73.54
				Tanh	0.02	0.09	1639	72.17
				ReLU	0.09	0.76	43	73.53
	PLSR-ANN	5	5	Logistic	0.08	0.08	522	72.16
				Tanh	0.04	0.05	421	72.09
				ReLU	0.09	0.77	47	73.53
Uttar Dinajpur	SR-ANN	3	1	Logistic	0.06	0.07	663	12.10
				Tanh	0.03	0.05	900	12.10
				ReLU	0.09	0.78	28	13.56
	PCA-ANN	5	2	Logistic	0.06	0.23	621	30.47
				Tanh	0.07	0.13	4439	30.26
				ReLU	0.09	0.78	43	31.57
	PLSR-ANN	6	2	Logistic	0.06	0.04	2200	34.06
				Tanh	0.02	0.03	1827	34.06
				ReLU	0.09	0.78	42	35.57

The learning rate of Tanh is also 0.07 and it converged to SSE of 0.12 in 1610 steps. The decision is to be in the direction of minimizing the error even with a comparatively larger learning rate (Kandil *et al* 1993). Since the SSE and AIC values is also low in logistic function, the logistics function is found to perform better in PCA-ANN. For the PLSR-ANN model, Tanh activation function is converged to the lowest possible SSE of 0.02 with a learning rate of 0.02 in 868 steps. Further, the AIC in Tanh function is also low. Hence, the Tanh activation function is found to perform better in the PLSR-ANN model.

4.4.2.1 b) Jalpaiguri

The RMSE-CV is low for five hidden layer neurons for SR-ANN and PLSR-ANN models and it is two for the PCA-ANN model. The number of input neurons are five for all three models.

In SR-ANN model, Tanh activation function is converged to the lowest possible SSE of 0.01 with a learning rate of 0.03 in 795 steps. Similarly, the Tanh activation function is converged to the lowest possible SSE of 0.02 with the learning rate of 0.03 in 360 steps in PLSR-ANN. Further, AIC in the Tanh function is also comparatively low in both SR-ANN and PLSR-ANN. Hence, the Tanh activation function is found to perform better in SR-ANN and PLSR-ANN models. The main advantage of Tanh activation function is that its output is zero-centered which makes the algorithm to learn faster (Nwankpa *et al* 2018). In PCA-ANN, logistic activation function is converged to the lowest possible SSE of 0.17 with a learning rate of 0.07 in 3583 steps. Further, AIC in logistic function is also comparatively low. Hence, the logistic activation function is found to perform better in the PCA-ANN model.

4.4.2.1 c) Malda

The RMSE-CV is low for five hidden layer neurons for all three ANN models. The number of input neurons is also five for all three models. The Tanh activation function is converged to the lowest possible SSE of 0.06, 0.09 and 0.05 with the learning rate of 0.02, 0.02 and 0.04 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. Further, the AIC in the Tanh function is low in all three ANN models. Hence, the Tanh activation function is found to perform better. The Tanh activation function can be used as an effective nonlinear function in the hidden layer (Feng and Lu 2019).

4.4.2.1 d) Uttar Dinajpur

The RMSE-CV is low for two hidden layer neurons for PCA-ANN and PLSR-ANN models and it is one for SR-NN model. The number of input neurons are three, five and six for SR-ANN, PCA-ANN and PLSR-ANN models respectively. The Tanh activation function is converged to the lowest possible SSE of 0.05, 0.13 and 0.03 with the learning rate of 0.03, 0.07 and 0.02 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. Further, the AIC in Tanh function is comparatively low in all three ANN models. Hence, the Tanh activation function is found to perform better.

It can be observed that the number of hidden layer neurons are less than or equal to the number of input layer neurons. It is in agreement with the results of Silaban and Zarlis (2017). More number of hidden layer neurons cause slow learning as well as overfitting issues and a smaller number of neurons may be insufficient. The ANN model with the optimum number of hidden layer neurons performs better (Meerasri *et al* 2022). The optimum learning rate ranges between 0.02 and 0.10. The optimum learning rate are in agreement with the results of Igiri *et al* (2015). The ANN models fitted by optimizing critical aspects such as learning rate and the number of hidden layer neurons increases the predictive power of the crop yield prediction model (Kaul *et al* 2005).

4.4.2.2 Wheat

The results of SR-ANN, PCA-ANN and PLSR-ANN models based on the backpropagation algorithm using three different activation functions for wheat crop of all four districts are given in Table 4.34.

4.4.2.2 a) Cooch Behar

The RMSE-CV is low for two, six and four hidden layer neurons for SR-ANN, PCA-ANN and PLSR-ANN models respectively. The number of input neurons are five for PCA-ANN and PLSR-ANN models and it is three for SR-ANN.

In SR-ANN model, Tanh activation function is converged to the lowest possible SSE of 0.14 with a learning rate of 0.02 in 577 iterations. The logistic function is also converged to SSE of 0.14 with a learning rate of 0.09 and it took comparatively many iterations to converge. The AIC value is also low in Tanh. Hence the Tanh function is found to perform better in SR-ANN. In PCA-ANN, Tanh activation function is converged to the lowest possible SSE of 0.05 with the learning rate of 0.02. Since the SSE and AIC value is also low in Tanh function, the Tanh function is found perform better in PCA-ANN. For PLSR-ANN model, both logistics and Tanh activation function is converged to the lowest possible SSE of 0.06. But Tanh function converged with comparatively lesser learning rate and iterations. Further, AIC in Tanh function is also low. Hence, Tanh activation function is found to perform better in PLSR-ANN model.

Table 4.34 Summary of ANN models using three activation functions for Wheat crop in four districts

District	Model	Number of Input Neurons	Number of Hidden layer Neurons	Activation Function	Learning Rate	SSE	Steps	AIC
Cooch Behar	SR-ANN	3	2	Logistic	0.09	0.14	1540	72.29
				Tanh	0.02	0.14	577	72.27
				ReLU	0.09	0.56	52	73.11
	PCA-ANN	5	6	Logistic	0.09	0.54	48	87.09
				Tanh	0.02	0.05	1691	86.11
				ReLU	0.09	0.56	47	87.11
	PLSR-ANN	5	4	Logistic	0.06	0.06	1739	58.14
				Tanh	0.02	0.06	1166	58.11
				ReLU	0.09	0.56	53	59.11
Jalpaiguri	SR-ANN	4	5	Logistic	0.05	0.05	1178	62.09
				Tanh	0.03	0.02	629	62.05
				ReLU	0.09	0.46	45	62.93
	PCA-ANN	6	8	Logistic	0.09	0.07	1845	130.15
				Tanh	0.02	0.03	1351	130.06
				ReLU	0.09	0.46	58	130.93
	PLSR-ANN	4	8	Logistic	0.09	0.47	59	98.93
				Tanh	0.01	0.04	1506	98.08
				ReLU	0.09	0.46	57	98.93
Malda	SR-ANN	3	5	Logistic	0.06	0.06	1442	52.11
				Tanh	0.03	0.03	5004	52.06
				ReLU	0.09	0.68	49	53.36
	PCA-ANN	5	6	Logistic	0.05	0.09	2003	86.17
				Tanh	0.01	0.03	7065	86.06
				ReLU	0.09	0.68	48	87.36
	PLSR-ANN	7	5	Logistic	0.05	0.04	1948	92.07
				Tanh	0.02	0.02	1667	92.05
				ReLU	0.09	0.68	46	93.36
Uttar Dinajpur	SR-ANN	3	5	Logistic	0.10	0.11	546	52.22
				Tanh	0.03	0.06	701	52.12
				ReLU	0.10	0.68	29	53.36
	PCA-ANN	5	3	Logistic	0.10	0.68	191	45.37
				Tanh	0.04	0.05	1007	44.10
				ReLU	0.10	0.68	25	45.36
	PLSR-ANN	4	3	Logistic	0.10	0.68	31	39.36
				Tanh	0.04	0.01	33634	38.02
				ReLU	0.10	0.68	34	39.36

4.4.2.2 b) Jalpaiguri

The number of input neurons is four for SR-ANN and PLSR-ANN models and it is six for PCA-ANN model. The RMSE-CV is low for eight hidden layer neurons for PCA-ANN and PLSR-ANN models and it is five for SR-NN model. The Tanh activation function is converged to the lowest possible SSE of 0.02, 0.03 and 0.04 with the learning rate of 0.03, 0.02 and 0.01 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. Further, the AIC in Tanh function is comparatively low in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.2 c) Malda

The number of input neurons are three, five and seven for SR-ANN, PCA-ANN and PLSR-ANN models respectively. The RMSE-CV is low for five hidden layer neurons for SR-ANN and PLSR-ANN models and it is six for PCA-ANN model. The Tanh activation function is converged to lowest possible SSE of 0.03, 0.03 and 0.02 with the learning rate of 0.03, 0.01 and 0.02 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. Further, the AIC in Tanh function is comparatively low in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.2 d) Uttar Dinajpur

The number of input neurons are three, five and four for SR-ANN, PCA-ANN and PLSR-ANN models respectively. The RMSE-CV is low for five hidden layer neurons for SR-ANN and it is three for PCA-ANN and PLSR-ANN models. The Tanh activation function is converged to lowest possible SSE of 0.06, 0.05 and 0.01 with the learning rate of 0.03, 0.04 and 0.04 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. Further, the AIC in Tanh function is low comparatively in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.3 Potato

The results of SR-ANN, PCA-ANN and PLSR-ANN models based on the backpropagation algorithm using three different activation functions for potato crop of all the four districts are given in Table 4.35.

4.4.2.3 a) Cooch Behar

The RMSE-CV is low for four hidden layer neurons for SR-ANN and PCA-ANN models and it is one for PLSR-ANN model. The number of input neurons are four, five and one for SR-ANN, PCA-ANN and PLSR-ANN respectively. The Tanh activation function is converged to the lowest possible SSE of 0.03, 0.03 and 0.25 with the learning rate of 0.01, 0.01 and 0.07 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. The Tanh function has taken many steps to converge but converge to the lowest possible error. Further, the AIC in Tanh function is comparatively low in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.3 b) Jalpaiguri

The number of input neurons are five for SR-ANN and PLSR-ANN models and it is four for PCA-ANN. The optimum number of hidden layer neurons are four for SR-ANN and PCA-ANN models and it is two for PLSR-ANN model. It can be observed in PCA-ANN that the Tanh function converged to SSE of zero with very small learning rate. Due to the smaller learning rate, the algorithm has taken many steps to converge. For SR-ANN, and PLSR-ANN models, the Tanh activation function is converged to lowest possible error of 0.05 with a learning rate of 0.03. The AIC in Tanh function is low in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.3 c) Malda

The number of input neurons are five for SR-ANN and PLSR-ANN models and it is four for PCA-ANN. The optimum number of hidden layer neurons are three, four and two for SR-ANN PCA-ANN and PLSR-ANN models respectively. The Tanh activation function is converged to the lowest possible SSE of 0.06, 0.17 and 0.04 with the learning rate of 0.03, 0.01 and 0.07 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. The Tanh function has taken many steps to converge but converge to the lowest possible error. Further, the AIC in Tanh function is comparatively low in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.3 d) Uttar Dinajpur

The number of input neurons is three for PCA-ANN and PLSR-ANN models and it is two for SR-ANN. The optimum number of hidden layer neurons are four for SR-ANN and PLSR-ANN models and it is five for PCA-ANN model. In SR-ANN model, ReLU activation function is converged to lowest possible error SSE of 0.03 with a learning rate of 0.05 in 3408 steps. The ReLU function has taken many steps to converge due to smaller learning rate. The AIC of ReLU function is also low. Hence the ReLU function is found to perform better in SR-ANN. The Tanh activation function is converged to the lowest possible error of 0.08 and 0.09 with the learning rate of 0.06 and 0.04 PCA-ANN and PLSR-ANN models respectively. Further, the AIC in Tanh function is low in all three ANN models. Hence, the Tanh activation function is found to perform better.

Table 4.35 Summary of ANN models using three activation functions for Potato crop in four districts

District	Model	Number of Input Neurons	Number of Hidden layer Neurons	Activation Function	Learning Rate	SSE	Steps	AIC
Cooch Behar	SR-ANN	4	4	Logistic	0.08	0.83	163	51.65
				Tanh	0.01	0.03	7906	50.06
				ReLU	0.09	0.82	43	51.63
	PCA-ANN	5	4	Logistic	0.08	0.24	493	58.47
				Tanh	0.01	0.03	3610	58.06
				ReLU	0.09	0.82	55	59.63
	PLSR-ANN	1	1	Logistic	0.08	0.26	543	8.52
				Tanh	0.07	0.25	2929	8.51
				ReLU	0.09	0.82	35	9.63
Jalpaiguri	SR-ANN	5	4	Logistic	0.09	0.79	50	59.57
				Tanh	0.03	0.05	1250	58.11
				ReLU	0.09	0.79	48	59.57
	PCA-ANN	5	4	Logistic	0.08	0.15	834	58.30
				Tanh	0.01	0.00	47951	58.00
				ReLU	0.09	0.79	52	59.57
	PLSR-ANN	4	2	Logistic	0.09	0.79	49	51.57
				Tanh	0.03	0.05	1075	50.09
				ReLU	0.09	0.79	44	51.57
Malda	SR-ANN	5	3	Logistic	0.07	0.07	5544	44.13
				Tanh	0.03	0.06	2096	44.13
				ReLU	0.09	1.02	42	46.04
	PCA-ANN	4	4	Logistic	0.09	1.03	49	52.05
				Tanh	0.01	0.17	6014	50.33
				ReLU	0.09	1.02	44	52.04
	PLSR-ANN	5	2	Logistic	0.06	0.08	3145	30.17
				Tanh	0.07	0.04	14105	30.08
				ReLU	0.09	1.02	42	32.04
Uttar Dinajpur	SR-ANN	4	2	Logistic	0.07	0.07	2754	26.15
				Tanh	0.06	0.07	315	26.13
				ReLU	0.03	0.05	3408	26.10
	PCA-ANN	5	3	Logistic	0.07	0.12	2883	44.24
				Tanh	0.06	0.08	2734	44.17
				ReLU	0.05	0.12	6598	44.23
	PLSR-ANN	4	3	Logistic	0.09	1.03	49	40.06
				Tanh	0.04	0.09	755	38.18
				ReLU	0.09	1.03	55	40.06

4.4.2.4 Jute

The results of SR-ANN, PCA-ANN and PLSR-ANN models based on the backpropagation algorithm using three different activation functions for jute crop of all the four districts are given in Table 4.36.

4.4.2.4 a) Cooch Behar

The number of input neurons are five for SR-ANN and it is four in PCA-ANN and PLSR-ANN models. The RMSE-CV is low for two hidden layer neurons for SR-ANN and PLSR-ANN models and it is one for PCA-ANN model. The Tanh activation function is converged to the lowest possible error of 0.04, 0.44 and 0.07 with the learning rate of 0.03, 0.09 and 0.07 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. The Tanh function has taken many steps to converge but converge to the lowest possible error. Further, the AIC in Tanh function is comparatively low in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.4 b) Jalpaiguri

The number of input neurons are four, five and six for SR-ANN, PCA-ANN and PLSR-ANN models respectively. The optimum number of hidden layer neurons are four for PCA-ANN and PLSR-ANN models and it is two for SR-ANN model. The Tanh activation function is converged to the lowest possible error of 0.03 and 0.04 with the learning rate of 0.05 and 0.02 in SR-ANN and PCA-ANN models respectively. Further, the AIC in Tanh function is comparatively low in SR-ANN and PCA-ANN. Hence, the Tanh activation function is found to perform better in SR-ANN and PCA-ANN models. In PLSR-ANN, both logistic and Tanh function converged to the lowest error of 0.04. But Tanh function has taken many steps to converge due to small learning rate. The AIC of logistic function is comparatively low. Hence, the PLSR-ANN model with logistic activation function is found to perform better.

4.4.2.4 c) Malda

The number of input neurons are three, five and four for SR-ANN, PCA-ANN and PLSR-ANN models respectively. The optimum number of hidden layer neurons are two, four and three for SR-ANN, PCA-ANN and PLSR-ANN models respectively. It can be observed in PCA-ANN that the Tanh function converged to SSE of zero with a learning rate of 0.03. Due to the smaller learning rate, the algorithm has taken many steps to converge. For SR-ANN, and PLSR-ANN models, the Tanh activation function is converged to the lowest possible error of 0.07 and 0.04 with a learning rate of 0.06 and 0.04 respectively. The AIC in Tanh function is low in all three ANN models. Hence, the Tanh activation function is found to perform better.

Table 4.36 Summary of ANN models using three activation functions for Jute crop in four districts

District	Model	Number of Input Neurons	Number of Hidden layer Neurons	Activation Function	Learning Rate	SSE	Steps	AIC
Cooch Behar	SR-ANN	5	2	Logistic	0.09	1.02	47	32.04
				Tanh	0.03	0.04	5213	30.08
				ReLU	0.09	1.02	47	32.04
	PCA-ANN	4	1	Logistic	0.07	0.45	966	14.90
				Tanh	0.09	0.44	752	14.88
				ReLU	0.07	0.88	73	15.75
	PLSR-ANN	4	2	Logistic	0.08	0.07	283	26.22
				Tanh	0.07	0.06	398	26.11
				ReLU	0.09	0.88	77	27.75
Jalpaiguri	SR-ANN	4	2	Logistic	0.08	0.04	505	26.07
				Tanh	0.05	0.03	452	26.05
				ReLU	0.09	0.70	39	27.40
	PCA-ANN	5	4	Logistic	0.09	0.12	95231	58.24
				Tanh	0.02	0.04	3214	58.07
				ReLU	0.09	0.70	52	59.40
	PLSR-ANN	6	4	Logistic	0.08	0.04	530	66.07
				Tanh	0.03	0.04	1369	66.08
				ReLU	0.09	0.70	40	67.40
Malda	SR-ANN	3	2	Logistic	0.08	0.08	1628	22.23
				Tanh	0.06	0.07	317	22.13
				ReLU	0.09	0.94	39	23.87
	PCA-ANN	5	4	Logistic	0.09	0.92	48	59.85
				Tanh	0.03	0.00	46406	58.00
				ReLU	0.09	0.94	54	59.87
	PLSR-ANN	4	3	Logistic	0.08	0.07	9747	38.14
				Tanh	0.04	0.04	24604	38.08
				ReLU	0.09	0.94	51	39.87
Uttar Dinajpur	SR-ANN	4	3	Logistic	0.09	0.01	30902	38.02
				Tanh	0.05	0.02	135	38.04
				ReLU	0.03	0.02	1597	38.04
	PCA-ANN	4	3	Logistic	0.09	0.58	49	39.17
				Tanh	0.04	0.14	271	38.28
				ReLU	0.09	0.58	53	39.16
	PLSR-ANN	5	1	Logistic	0.07	0.03	820	16.05
				Tanh	0.07	0.02	352	16.03
				ReLU	0.09	0.58	40	17.16

4.4.2.4 d) Uttar Dinajpur

The number of input neurons are four for SR-ANN and PCA-ANN models and it is five for PLSR-ANN. The optimum number of hidden layer neurons are three for SR-ANN and PCA-ANN models and it is one for PLSR-ANN model. In SR-ANN model, logistic activation function is converged to lowest possible error SSE of 0.01 with a learning rate of 0.09. The logistic function has taken many steps to converge to lowest possible error. The AIC of logistic function is also low. Hence, the logistic function is found to perform better in SR-ANN. The Tanh activation function is converged to lowest possible error of 0.14 and 0.02 with the learning rate of 0.04 and 0.07 in PCA-ANN and PLSR-ANN models respectively. Further, the AIC in Tanh function is comparatively low in PCA-ANN and PLSR-ANN. Hence, the PCA-ANN and PLSR-ANN models with Tanh activation function is found to perform better.

4.4.2.5 Rapeseed-Mustard

The results of SR-ANN, PCA-ANN and PLSR-ANN models based on the backpropagation algorithm using three different activation functions for rapeseed-mustard crop of all four districts is given in Table 4.37.

4.4.2.5 a) Cooch Behar

The number of input neurons are three, five and two for SR-ANN, PCA-ANN for PLSR-ANN models respectively. The RMSE-CV is low for three hidden layer neurons for SR-ANN and PCA-ANN models and it is four for PLSR-ANN model. The Tanh activation function is converged to the lowest possible error of 0.14 and 0.01 with the learning rate of 0.03 and 0.05 in SR-ANN and PCA-ANN models respectively. The Tanh activation function in PCA has taken many steps to converge very lowest error of 0.01. The AIC in Tanh function is low in SR-ANN and PCA-ANN. Hence the SR-ANN and PCA-ANN models with Tanh activation function is found to perform better. In PLSR-ANN model, logistic activation function is converged to lowest possible error of 0.07 with a learning rate of 0.07. The AIC of logistic function is also low. Hence, the logistic function is found to perform better in PLSR-ANN.

4.4.2.5 b) Jalpaiguri

The number of input neurons are five for SR-ANN and PCA-ANN models and it is four for PLSR-ANN. The optimum number of hidden layer neurons is three in SR-ANN and PLSR-ANN models and it is five for PCA-ANN model. The Tanh activation function is converged to the lowest possible SSE of 0.05, 0.04 and 0.06 with the learning rate of 0.04, 0.03 and 0.03 in SR-ANN, PCA-ANN and PLSR-ANN models respectively. The Tanh function has taken many steps to converge but converged to the lowest possible error. Further, the AIC in Tanh function is comparatively low in all three ANN models. Hence, the Tanh activation function is found to perform better.

4.4.2.5 c) Malda

The number of input neurons are four for SR-ANN and PLSR-ANN models and it is five for PCA-ANN. The optimum number of hidden layer neurons are two, five and three in SR-ANN, PCA-ANN and PLSR-ANN models respectively. In SR-ANN model, logistic activation function is converged to the lowest possible error of 0.04 with a learning rate of 0.07. The logistic function has taken many steps to converge but converged to the lowest possible error. The AIC of the logistic function is also comparatively low. Hence the logistic function is found to perform better in SR-ANN. In PCA-ANN model, Tanh activation function is converged to lowest possible SSE of 0.07 with the learning rate of 0.03 in 358 steps. Similarly, the Tanh activation function is converged to lowest possible error of 0.02 with a learning rate of 0.03 in PLSR-ANN. The Tanh function in PLSR-ANN has taken many steps to converge but converged to the lowest possible error of 0.02. Further, AIC in Tanh function is also comparatively low in both PCA-ANN and PLSR-ANN. Hence, SR-ANN and PLSR-ANN models with Tanh activation function is found to perform better.

4.4.2.5 d) Uttar Dinajpur

The number of input neurons is three, five and one for SR-ANN, PCA-ANN and PLSR-ANN models respectively. The optimum number of hidden layer neurons are three in SR-ANN and it is two in PCA-ANN and PLSR-ANN models. The Tanh activation function is converged to the lowest possible error of 0.12 and 0.01 with the learning rate of 0.03 and 0.07 in SR-ANN and PCA-ANN models respectively. Further, the AIC in Tanh function is comparatively low. Hence SR-ANN and PCA-ANN models with Tanh activation function is found to perform better. In PLSR-ANN model, both logistic and Tanh activation function is converged to lowest possible error of 0.07 with a learning rate of 0.17. But the Tanh function has taken comparatively more steps to converge to the same error of 0.17. The AIC of logistic function is also comparatively lesser than Tanh. Hence, the PLSR-ANN model with logistic activation function is found to perform better.

Table 4.37 Summary of ANN models using three activation functions for Rapeseed-mustard crop in four districts

District	Model	Number of Input Neurons	Number of Hidden layer Neurons	Activation Function	Learning Rate	SSE	Steps	AIC
Cooch Behar	SR-ANN	3	3	Logistic	0.06	0.17	2215	32.35
				Tanh	0.03	0.14	3329	32.27
				ReLU	0.09	0.88	50	33.75
	PCA-ANN	5	3	Logistic	0.09	0.88	47	45.75
				Tanh	0.05	0.01	85006	44.03
				ReLU	0.09	0.84	78	45.69
	PLSR-ANN	2	4	Logistic	0.07	0.07	5414	34.34
				Tanh	0.04	0.19	237	34.38
				ReLU	0.09	0.88	41	35.75
Jalpaiguri	SR-ANN	5	3	Logistic	0.06	0.06	974	44.11
				Tanh	0.04	0.05	600	44.09
				ReLU	0.09	0.62	42	45.25
	PCA-ANN	5	5	Logistic	0.09	0.64	47	73.28
				Tanh	0.03	0.04	1237	72.07
				ReLU	0.09	0.60	136	73.19
	PLSR-ANN	4	3	Logistic	0.09	0.62	49	39.25
				Tanh	0.03	0.06	1031	38.11
				ReLU	0.09	0.62	54	39.25
Malda	SR-ANN	4	2	Logistic	0.07	0.04	8068	26.09
				Tanh	0.06	0.05	180	26.10
				ReLU	0.09	0.52	41	27.05
	PCA-ANN	5	5	Logistic	0.09	0.53	239	73.07
				Tanh	0.03	0.07	358	72.14
				ReLU	0.09	0.52	43	73.05
	PLSR-ANN	4	3	Logistic	0.09	0.52	48	39.05
				Tanh	0.03	0.02	18764	38.04
				ReLU	0.09	0.52	55	39.05
Uttar Dinajpur	SR-ANN	3	3	Logistic	0.05	0.16	452	32.31
				Tanh	0.03	0.12	832	32.23
				ReLU	0.09	0.78	48	33.56
	PCA-ANN	5	2	Logistic	0.07	0.17	521	30.34
				Tanh	0.07	0.01	2218	30.20
				ReLU	0.09	0.78	41	31.56
	PLSR-ANN	1	2	Logistic	0.06	0.17	233	14.34
				Tanh	0.04	0.17	698	14.35
				ReLU	0.09	0.78	42	15.56

4.4.3 Support Vector Regression (SVR) Model

Three Support Vector Regression (SVR) models have been fitted to each crop in each district by taking indices selected from stepwise regression as input (SR-SVR), principal components as input (PCA-SVR) and PLSR components as input (PLSR-SVR). The performance of SVR model under a Linear kernel as well as two nonlinear kernels *viz.* Polynomial and Radial Basis Function (RBF) have been evaluated. Linear kernel has only two hyperparameters, cost and epsilon. RBF kernel has an additional parameter gamma. The range of cost (C) parameter is set as 1 to 32. The range of epsilon (ε) and gamma (γ) are set from 0 to 1 with 0.01 interval. The degree of polynomial (d) is the fourth parameter in the polynomial kernel which is tuned from 1 to 10. The best combination of these hyperparameters is selected for each kernels using grid search algorithm based on 10-fold cross validation technique.

4.4.3.1 Rice

The results of SR-SVR, PCA-SVR and PLSR-SVR models using three different kernel functions for rice crop of all four districts are given in Table 4.38.

4.4.3.1 a) Cooch Behar

The optimum cost (C) is the lowest level of one for linear kernel of all three models. Similarly, the optimum width of the hyperplane (ε) is also at lowest level of 0.01 for linear kernel of all three models. In the case of polynomial kernel, the degree of polynomial is one for all the three models which is similar to linear kernel. But the cost of polynomial kernel increased to 4 in SR-SVR and PLSR-SVR and it is highest level of 32 in PCA-SVR. The width of the margin (ε) is zero in polynomial kernel of all the three models due to which the number of support vectors are high level of 20. The low margin (ε) leads to many support vectors and high ε leads to a reduction in number of support vectors (Alrefaee *et al* 2022). The model with huge number of support vectors may fit to the training data well. However, the generalized performance of the model for new data during validation is generally poor due to the overfitting issue (Nanda *et al* 2018). The gamma (γ) which is the symbol of curvature of kthe ernel is greater than zero. The γ=1 in PCA-SVR indicates the existence of more nonlinearity in the model. In RBF kernel, the optimum cost is four in SR-SVR and it is eight for PCA-SVR and PLSR-SVR models. The number of support vectors are high level of 20 in RBF function of in SR-SVR and PLSR-SVR since the width of the hyperplane margin (ε) is zero. The ε value of 0.02 leads to a considerable reduction in the number of support vectors in PCA-SVR model. It can be observed that a small increase in the ε in linear kernel makes more reduction in support vectors than other two kernels. On the whole, the RBF kernel with optimum choices of hyperparameter achieved the lowest RMSE in all the three models. Similar result is obtained by Guo *et al* 2020.

4.4.3.1 b) Jalpaiguri

The optimum C and ε is the lowest level of 1 and 0.1 respectively for linear kernel of all the three models. The degree of polynomial of polynomial kernel is one for all the three models. But the cost of polynomial kernel increased to 4 in SR-SVR and PLSR-SVR and it is the highest level of 32 in PCA-SVR. The ε is zero in polynomial kernel of all three models due to which the number of support vectors are high level of 20. The non-zero γ of both polynomial and RBF indicates the presence of nonlinearity. In RBF kernel, the optimum cost is four in SR-SVR and PLSR-SVR models and it is eight for PCA-SVR model. The width of the hyperplane margin (ε) is increased to 0.1 in RBF kernel of SR-SVR due to which there is a reduction in the number of support vectors. Further increase in the ε of RBF to 15 leads to a further reduction in the support vectors. The RBF kernel with optimum choices of hyperparameter performed better with the lowest RMSE in all three models. On the whole, the RBF kernel with PLSR components as input achieved the lowest RMSE.

4.4.3.1 c) Malda

The optimum C and ε is lowest levels of 1 and 0.1 respectively for the linear kernel of all three models. The degree of polynomial of polynomial kernel is one for all the three models. The ε of 0.1 in polynomial kernel leads to a reduction in the number of support vectors. The non-zero γ of both polynomial and RBF indicates the presence of nonlinearity. In RBF kernel, the optimum cost is four in all three models. The width of the hyperplane margin (ε) is zero in RBF kernel of all the models due to which the number of support vectors is highest level of 20. The RBF kernel with optimum choices of hyperparameter performed better with the lowest RMSE in all the three models. On the whole, the RBF kernel with PLSR components as input achieved the lowest RMSE.

4.4.3.1 d) Uttar Dinajpur

The optimum C and ε is the lowest level of 1 and 0.1 respectively for linear kernel of all three models. The degree of polynomial of polynomial kernel is one for all the three models. There is a reduction in the number of support vectors in polynomial kernel as ε is greater than zero. The γ of polynomial kernel is greater than RBF which indicates the presence of more nonlinearity in the former than latter. In the RBF kernel, the optimum cost is four in all three models. The number of support vectors in RBF is reduced as the ε is increased from zero but the reduction is lesser than the other two kernels. The RBF kernel with optimum choices of hyperparameter performed better with the lowest RMSE in all the three models. On a whole, the RBF kernel with PLSR components as input achieved the lowest RMSE.

Table 4.38 Summary of Support Vector Regression (SVR) model using three kernels for Rice crop in four districts

Model	Kernel Function	Cost (C)	Epsilon (ε)	Gamma (γ)	Degree (d)	Number of Support Vectors	RMSE
Cooch Behar							
SR-SVR	Linear	1	0.10	N/A	N/A	12	63.36
	Polynomial	4	0.00	0.20	1	20	64.08
	RBF	4	0.00	0.10	N/A	20	18.85
PCA-SVR	Linear	1	0.10	N/A	N/A	15	206.09
	Polynomial	32	0.00	1.00	1	20	205.97
	RBF	8	0.02	0.20	N/A	17	91.69
PLSR-SVR	Linear	1	0.10	N/A	N/A	17	67.33
	Polynomial	4	0.00	0.10	1	20	63.05
	RBF	8	0.00	0.10	N/A	20	0.13
Jalpaiguri							
SR-SVR	Linear	1	0.10	N/A	N/A	9	53.35
	Polynomial	4	0.00	0.30	1	20	52.35
	RBF	4	0.10	0.10	N/A	15	43.73
PCA-SVR	Linear	1	0.10	N/A	N/A	17	213.54
	Polynomial	4	0.00	0.10	1	20	219.30
	RBF	8	0.50	0.10	N/A	12	202.53
PLSR-SVR	Linear	1	0.10	N/A	N/A	11	58.61
	Polynomial	16	0.10	0.60	1	11	59.11
	RBF	4	0.00	0.10	N/A	20	7.68
Malda							
SR-SVR	Linear	1	0.10	N/A	N/A	15	129.20
	Polynomial	4	0.10	0.10	1	14	130.29
	RBF	4	0.00	0.10	N/A	20	111.01
PCA-SVR	Linear	1	0.10	N/A	N/A	19	216.84
	Polynomial	32	0.10	0.70	1	19	218.90
	RBF	4	0.00	0.10	N/A	20	138.44
PLSR-SVR	Linear	1	0.10	N/A	N/A	13	117.17
	Polynomial	4	0.00	0.20	1	20	116.61
	RBF	4	0.00	0.10	N/A	20	95.3
Uttar Dinajpur							
SR-SVR	Linear	1	0.10	N/A	N/A	18	104.99
	Polynomial	8	0.20	0.80	1	13	104.47
	RBF	32	0.20	0.20	N/A	12	76.54
PCA-SVR	Linear	1	0.10	N/A	N/A	18	245.32
	Polynomial	4	0.30	0.10	1	12	238.87
	RBF	4	0.20	0.10	N/A	17	175.21
PLSR-SVR	Linear	1	0.10	N/A	N/A	18	86.35
	Polynomial	8	0.30	0.60	1	15	100.64
	RBF	16	0.10	0.10	N/A	17	41.69

The performance of RBF kernel is better than other kernels in terms of the stability of SVR model (Truong and Pham, 2021). The RBF kernel accompanied with parameter tuning reduces predictive error than polynomial and linear kernels (Guenther and Schonlau, 2016). The grid search algorithm is effective for optimizing the hyperparameters RBF kernel (Agustina *et al* 2018, Ngoc *et al* 2021).

4.4.3.2 Wheat

The results of SR-SVR, PCA-SVR and PLSR-SVR models using three different kernel functions for wheat crop of all four districts are given in Table 4.39.

4.4.3.2 a) Cooch Behar

The optimum cost (C) is the lowest level of one for linear kernel of all the three models. Similarly, the optimum width of the hyperplane (ε) is also at lowest level of 0.01 for the linear kernel of all three models. In the case of polynomial kernel, the degree of polynomial is one for all the three models. The width of the hyperplane (ε) of 0.3 in polynomial kernel of SR-SVR leads to a reduction in number of support vectors to 10. The ε of 0.6 in polynomial kernel in PCA-SVR leads to a further reduction in number of support vector to the level of 7. The ε is zero in RBF kernel of PCA-SVR due to which the number of support vectors is high level of 20. The ε of 0.1 in RBF kernel of SR-SVR leads to a reduction in support vectors. The ε of 0.4 in RBF of PLSR-SVR leads to further reduction in the number of support vectors to 11. It can be seen that the reduction in the number of support vectors for increase in ε is less in RBF. The γ=1 in PLSR-SVR indicates the existence of more nonlinearity in the model. The γ is 0.1 for the remaining models of polynomial as well as RBF kernel. The optimum cost is high level of 32 in polynomial kernel of PLSR-SVR followed by RBF of the same model. The RBF kernel with optimum choices of hyperparameter performed better with lowest RMSE in all three models. On the whole, the RBF kernel with PLSR components as input achieved the lowest RMSE.

4.4.3.2 b) Jalpaiguri

The optimum C and ε is the lowest level of 1 and 0.1 respectively for linear kernel of all three models. The ε of zero in both polynomial, as well as RBF kernels of SR-SVR, leads to many support vectors. The ε=1 in polynomial kernel of PCA-SVR, due to which there is a huge decrease in the number of support vectors to 4. It can be noted that the degree is 5 and the γ is 0.40 which indicates that there is a high nonlinearity in the polynomial kernel of PCA-SVR model. The cost is the highest level of 32 in RBF kernel of PCA-SVR and it is four for both the non-linear kernels of other models. The RMSE in PCA-SVR model with RBF kernel is at the lowest level of 0.14. The RMSE of SR-SVR and PLSR-SVR also low in RBF kernel. The RBF kernel with optimum choices of hyperparameter performed better with lowest RMSE in all the three models.

Table 4.39 Summary of Support Vector Regression (SVR) model using three kernels for Wheat crop in four districts

Model	Kernel Function	Cost (C)	Epsilon (ε)	Gamma (γ)	Degree (d)	Number of Support Vectors	RMSE
Cooch Behar							
SR-SVR	Linear	1	0.10	N/A	N/A	17	188.29
	Polynomial	4	0.30	0.10	1	10	197.35
	RBF	4	0.10	0.10	N/A	18	178.17
PCA-SVR	Linear	1	0.10	N/A	N/A	18	225.15
	Polynomial	4	0.60	0.10	1	7	213.08
	RBF	8	0.00	0.10	N/A	20	77.43
PLSR-SVR	Linear	1	0.10	N/A	N/A	16	156.75
	Polynomial	32	0.00	1.00	1	20	171.66
	RBF	16	0.40	0.10	N/A	11	118.16
Jalpaiguri							
SR-SVR	Linear	1	0.10	N/A	N/A	14	156.49
	Polynomial	4	0.00	0.10	1	20	170.83
	RBF	4	0.00	0.10	N/A	20	92.18
PCA-SVR	Linear	1	0.10	N/A	N/A	18	191.56
	Polynomial	4	1.00	0.40	5	4	286.54
	RBF	32	0.00	0.01	N/A	20	0.14
PLSR-SVR	Linear	1	0.10	N/A	N/A	12	140.27
	Polynomial	4	0.00	0.10	1	20	148.53
	RBF	4	0.10	0.10	N/A	16	86.76
Malda							
SR-SVR	Linear	1	0.10	N/A	N/A	18	126.01
	Polynomial	4	0.70	0.30	1	4	126.59
	RBF	8	0.00	0.10	N/A	20	63.12
PCA-SVR	Linear	1	0.10	N/A	N/A	19	158.19
	Polynomial	16	0.70	0.70	1	7	155.10
	RBF	8	0.00	0.20	N/A	20	15.64
PLSR-SVR	Linear	1	0.10	N/A	N/A	17	82.90
	Polynomial	4	0.10	0.20	1	15	82.23
	RBF	4	0.00	0.10	N/A	20	11.58
Uttar Dinajpur							
SR-SVR	Linear	1	0.10	N/A	N/A	17	114.79
	Polynomial	8	0.20	0.20	3	13	101.74
	RBF	32	0.40	0.01	N/A	10	95.34
PCA-SVR	Linear	1	0.10	N/A	N/A	14	118.36
	Polynomial	4	0.20	0.20	1	11	116.51
	RBF	4	0.30	0.10	N/A	11	86.90
PLSR-SVR	Linear	1	0.10	N/A	N/A	13	94.21
	Polynomial	4	0.60	0.50	1	4	106.74
	RBF	4	0.10	0.10	N/A	17	52.95

4.4.3.2 c) Malda

The optimum C and ε is the lowest level of 1 and 0.1 respectively for linear kernel of all the three models. The ε of 0.7 in the polynomial kernel of SR-SVR leads to the reduction in the number of support vectors to 4. The ε of polynomial kernel in PCA-SVR model is 0.7, the γ is 0.7 and the cost is high level of 16 which leads to 7 support vectors. The number of support vectors are 20 for RBF kernel of all the three models as the ε is zero. The RMSE is less in RBF kernels of all three models and it at the lowest level of 11.58 in PLSR-SVR.

4.4.3.2 d) Uttar Dinajpur

The optimum C and ε is lowest level of 1 and 0.1 respectively for linear kernel of all three models. The ε is increased to 0.20 in polynomial kernel of SR-SVR model leading to the reduction in support vectors to 13. Further, decrease in ε to 0.40 in RBF kernel results in further reduction in support vectors to 10. It can be noted that the degree of polynomial kernel is 3 which indicates that there is a considerable nonlinearity in the model. It can be observed that the ε is 0.20 and 0.30 in polynomial and RBF kernel of PCA-SVR respectively which results in same number of support vectors of 11. The ε of 0.60 and γ of 0.50 result in a reduction in the number of support vectors to 4. The C is high in RBF kernel in SR-SVR model. The RMSE is lowest level of 52.95 in RBF kernel of PLSR-SVR model.

4.4.3.3 Potato

The results of SR-SVR, PCA-SVR and PLSR-SVR models using three different kernel functions for potato crop of all four districts are given in the Table 4.40.

4.4.3.3 a) Cooch Behar

The optimum C and ε is the lowest levels of 1 and 0.1 respectively for linear kernel of all the three models. The ε of 0.80 and γ of 0.50 in polynomial kernel of SR-SVR leads to the reduction in the number support vectors to 6. The degree of polynomial is one for all three models. The number of support vectors are 20 for RBF kernel in SR-SVR and PCA-SVR as the ε is zero. The ε of 0.30 in RBF in PLSR-SVR leads to 13 number of support vectors. There is a high level of non-linearity in polynomial kernel in PLSR-SVR as the γ is one. The RMSE is less in RBF kernels of all three models and it is lowest level of 0.22 in SR-SVR.

Table 4.40 Summary of Support Vector Regression (SVR) model using three kernels for Potato crop in four districts

Model	Kernel Function	Cost (C)	Epsilon (ε)	Gamma (γ)	Degree (d)	Number of Support Vectors	RMSE
Cooch Behar							
SR-SVR	Linear	1	0.10	N/A	N/A	16	3.50
	Polynomial	4	0.80	0.40	1	6	3.86
	RBF	32	0.00	0.10	N/A	20	0.22
PCA-SVR	Linear	1	0.10	N/A	N/A	17	4.08
	Polynomial	4	0.00	0.10	1	20	4.06
	RBF	4	0.00	0.30	N/A	20	0.61
PLSR-SVR	Linear	1	0.10	N/A	N/A	14	3.91
	Polynomial	32	0.00	1.00	1	20	3.91
	RBF	8	0.30	0.10	N/A	13	3.69
Jalpaiguri							
SR-SVR	Linear	1	0.10	N/A	N/A	16	1.44
	Polynomial	4	0.30	0.10	1	7	1.43
	RBF	4	0.00	0.10	N/A	20	0.53
PCA-SVR	Linear	1	0.10	N/A	N/A	16	2.94
	Polynomial	4	0.30	0.10	1	11	2.77
	RBF	8	0.30	0.10	N/A	14	1.73
PLSR-SVR	Linear	1	0.10	N/A	N/A	18	1.60
	Polynomial	4	0.20	1.00	1	13	1.54
	RBF	8	0.00	0.10	N/A	20	0.36
Malda							
SR-SVR	Linear	1	0.10	N/A	N/A	18	3.04
	Polynomial	4	0.30	0.80	1	12	3.03
	RBF	16	0.00	0.10	N/A	20	0.00
PCA-SVR	Linear	1	0.10	N/A	N/A	19	5.63
	Polynomial	4	0.70	0.30	1	7	6.06
	RBF	8	0.00	0.30	N/A	20	1.56
PLSR-SVR	Linear	1	0.10	N/A	N/A	17	3.06
	Polynomial	4	0.40	0.10	1	15	2.99
	RBF	8	0.00	0.10	N/A	20	1.11
Uttar Dinajpur							
SR-SVR	Linear	1	0.10	N/A	N/A	18	1.81
	Polynomial	8	0.00	0.60	1	20	1.87
	RBF	4	0.40	0.01	N/A	12	2.00
PCA-SVR	Linear	1	0.10	N/A	N/A	18	2.79
	Polynomial	4	0.60	0.10	1	6	2.79
	RBF	4	0.20	0.10	N/A	15	1.40
PLSR-SVR	Linear	1	0.10	N/A	N/A	18	1.81
	Polynomial	32	0.20	0.70	1	15	1.85
	RBF	4	0.20	0.10	N/A	15	1.36

4.4.3.3 b) Jalpaiguri

The ε of 0.30 and γ of 0.10 in polynomial of SR-SVR leads to 6 support vectors. The same level of ε and γ in polynomial and RBF kernel in PCA-SVR results in 11 and 14 number of support vectors respectively. The degree of polynomial is one for all the three models. The number of support vectors are 20 for RBF kernel of PLSR-SVR as the ε is zero and it is 13 in polynomial kernel since the ε is 0.20. The γ is one in polynomial kernel of PLSR-SVR model indicates high nonlinearity in the model. The RMSE is comparatively less in RBF of all the three models and it is the lowest level of 0.36 in PLSR-SVR model.

4.4.3.3 c) Malda

The number of support vectors are 20 for RBF kernel of all three models as the ε is zero. The ε of 0.30, 0.70 and 0.40 in the polynomial kernel of SR-SVR, PCA-ANN and PLSR-SVR models which results to 12, 7 and 15 support vectors. The γ is 0.80 in polynomial kernel of SR-SVR. The γ is 0.30 in polynomial and RBF kernels of PCA-SVR model. The RMSE is less in RBF kernel in all three models and it is lowest level of zero in SR-SVR model.

4.4.3.3 d) Uttar Dinajpur

The ε of 0.40 in RBF kernel of SR-SVR model results in 12 support vectors and the number of support vectors in both PCA-SVR and PLSR-SVR is 15 with ε of 0.20 and γ of 0.10. The ε of 0.60 in polynomial kernel results to the lowest number of support vectors of 6. The linear kernel with ε value of 0.1 and 18 number of support vectors results in comparatively lesser RMSE among the three kernels of SR-SVR model. The RMSE is less in RBF kernel in PCA-SVR and PLSR-SVR.

4.4.3.4 Jute

The results of SR-SVR, PCA-SVR and PLSR-SVR models using three different kernel functions for jute crop of all four districts are given in Table 4.41.

4.4.3.4 a) Cooch Behar

The ε of 0.10 in RBF kernel of SR-SVR model results in 17 support vectors. It can be noted that the optimum ε of RBF kernel in PCA-SVR model is at high level of 1.0 due to which the number of support vectors reduced to 8. But the linear kernel in PCA-SVR with ε of 0.1 and having 19 support vectors achieved less RMSE among the three kernels in PCA-SVR model. In PLSR-SVR model, the RBF kernel achieved comparatively lesser RMSE with ε and γ of 0.1. The lowest RMSE is obtained in RBF kernel in both SR-SVR and PLSR-SVR models.

Table 4.41 Summary of Support Vector Regression (SVR) model using three kernels for Jute crop in four districts

Model	Kernel Function	Cost (C)	Epsilon (ε)	Gamma (γ)	Degree (d)	Number of Support Vectors	RMSE
Cooch Behar							
SR-SVR	Linear	1	0.10	N/A	N/A	13	0.38
	Polynomial	4	0.00	0.10	1	20	0.40
	RBF	4	0.10	0.10	N/A	17	0.26
PCA-SVR	Linear	1	0.10	N/A	N/A	19	1.10
	Polynomial	4	0.00	0.10	3	20	1.16
	RBF	4	1.00	0.10	N/A	8	1.25
PLSR-SVR	Linear	1	0.10	N/A	N/A	16	0.45
	Polynomial	4	0.10	0.40	1	16	0.45
	RBF	4	0.10	0.10	N/A	18	0.29
Jalpaiguri							
SR-SVR	Linear	1	0.10	N/A	N/A	15	0.41
	Polynomial	4	0.10	0.10	1	13	0.41
	RBF	8	0.00	0.10	N/A	20	0.12
PCA-SVR	Linear	1	0.10	N/A	N/A	16	1.26
	Polynomial	4	0.70	0.10	3	4	1.52
	RBF	8	0.00	0.10	N/A	20	0.42
PLSR-SVR	Linear	1	0.10	N/A	N/A	17	0.43
	Polynomial	4	0.50	0.30	1	7	0.50
	RBF	8	0.10	0.20	N/A	15	0.38
Malda							
SR-SVR	Linear	1	0.10	N/A	N/A	15	0.73
	Polynomial	4	0.00	0.10	1	14	0.75
	RBF	4	0.10	0.30	N/A	11	0.62
PCA-SVR	Linear	1	0.10	N/A	N/A	17	1.36
	Polynomial	4	0.10	0.70	1	17	1.36
	RBF	8	0.10	0.80	N/A	9	1.35
PLSR-SVR	Linear	1	0.10	N/A	N/A	16	0.76
	Polynomial	4	0.20	0.10	1	12	0.74
	RBF	8	0.10	0.10	N/A	17	0.46
Uttar Dinajpur							
SR-SVR	Linear	1	0.10	N/A	N/A	14	0.40
	Polynomial	4	0.30	0.20	1	5	0.44
	RBF	16	0.00	0.10	N/A	20	0.29
PCA-SVR	Linear	1	0.10	N/A	N/A	20	1.12
	Polynomial	16	0.70	0.70	1	7	1.17
	RBF	16	0.00	0.10	N/A	20	0.24
PLSR-SVR	Linear	1	0.10	N/A	N/A	13	0.30
	Polynomial	4	0.10	0.40	1	13	0.30
	RBF	4	0.00	0.10	N/A	20	0.06

4.4.3.4 b) Jalpaiguri

The number of support vectors is 20 in RBF kernel of SR-SVR and PCA-SVR models since their ε is zero. The RBF kernel of PLSR-SVR model which has the ε of 0.10, γ of 0.20 and 15 support vectors achieved the lowest RMSE of 0.38 among the other two kernels. An increase in the ε value of the polynomial kernel of SR-SVR and PCA-SVR models leads to a drastic reduction in the number of support vectors but their RMSE is higher than the RBF kernel. On the whole, the SR-SVR model with RBF kernel achieved the lowest RMSE of 0.12 among all the kernels of all the three models.

4.4.3.4 c) Malda

In SR-SVR model, the RBF kernel with ε of 0.10, Cost of 4 and γ of 0.30 results in 11 support vectors and the RBF kernel achieved the lowest RMSE among other two kernels of the SR-SVR model. The RBF kernel which has ε of 0.10 and γ of 0.80 results in 9 support vectors and comparatively lower RMSE is obtained in RBF kernel of the PCA-SVR model. Similarly ,the RBF kernel of PLSR-SVR model achieved comparatively lesser RMSE with ε and γ of 0.1. The PLSR-SVR model with RBF kernel achieved the lowest RMSE of 0.46 among all the kernels of all three models.

4.4.3.4 d) Uttar Dinajpur

The RBF kernel results in the highest level of 20 support vectors since the ε is zero in all three models. The γ value of greater than zero indicates certain nonlinearity in the model. The same kernel achieved a comparatively lower RMSE among the other two kernels in all three models. There was a reduction in the number of support vectors in polynomial as well as linear kernels. But their RMSE is comparatively high. The PLSR-SVR model with RBF kernel achieved the lowest RMSE of 0.06 among all the kernels of all three models.

4.4.3.5 Rapeseed-Mustard

The results of SR-SVR, PCA-SVR and PLSR-SVR models using three different kernel functions for the rapeseed-mustard crop of all four districts are given in Table 4.42.

4.4.3.5 a) Cooch Behar

In SR-SVR model, the RBF kernel with ε of 0.40, Cost of 4 and γ of 0.10 results in 9 support vectors and the RBF kernel achieved the lowest RMSE among other two kernels of the SR-SVR model. The RBF kernel which has ε of 0.40 and γ of 0.10 results in 14 support vectors and comparatively lower RMSE is obtained in the RBF kernel of the PCA-SVR model. Similarly, the RBF kernel of PLSR-SVR has the optimum ε of 0.70 and γ of 0.1 which results in 9 support vectors. The RBF kernel achieved comparatively lesser RMSE. The PCA-SVR model with RBF kernel achieved the lowest RMSE of 55.51 among all the kernels of all the three models.

Table 4.42 Summary of Support Vector Regression (SVR) model using three kernels for Rapeseed-mustard crop in four districts

Model	Kernel Function	Cost (C)	Epsilon (ε)	Gamma (γ)	Degree (d)	Number of Support Vectors	RMSE
Cooch Behar							
SR-SVR	Linear	1	0.10	N/A	N/A	19	70.90
	Polynomial	4	0.70	0.20	1	4	71.10
	RBF	4	0.40	0.10	N/A	9	61.85
PCA-SVR	Linear	1	0.10	N/A	N/A	17	85.12
	Polynomial	4	0.70	0.40	1	8	84.75
	RBF	16	0.40	0.10	N/A	14	55.51
PLSR-SVR	Linear	1	0.10	N/A	N/A	17	74.97
	Polynomial	4	0.50	0.20	1	8	71.57
	RBF	32	0.70	0.10	N/A	5	69.84
Jalpaiguri							
SR-SVR	Linear	1	0.10	N/A	N/A	17	43.19
	Polynomial	4	0.00	0.10	1	20	43.15
	RBF	4	0.00	0.10	N/A	20	25.13
PCA-SVR	Linear	1	0.10	N/A	N/A	17	64.23
	Polynomial	4	0.00	0.10	1	20	64.12
	RBF	4	0.00	0.10	N/A	20	23.44
PLSR-SVR	Linear	1	0.10	N/A	N/A	14	41.39
	Polynomial	4	0.10	0.20	1	14	41.39
	RBF	32	0.00	0.10	N/A	20	0.04
Malda							
SR-SVR	Linear	1	0.10	N/A	N/A	16	60.17
	Polynomial	32	0.30	1.00	1	9	57.82
	RBF	32	0.00	0.10	N/A	20	23.10
PCA-SVR	Linear	1	0.10	N/A	N/A	18	72.20
	Polynomial	4	0.50	0.70	1	6	74.03
	RBF	4	0.50	0.10	N/A	10	78.29
PLSR-SVR	Linear	1	0.10	N/A	N/A	18	50.24
	Polynomial	4	0.30	0.40	1	8	52.85
	RBF	8	0.10	0.10	N/A	17	25.93
Uttar Dinajpur							
SR-SVR	Linear	1	0.10	N/A	N/A	16	97.22
	Polynomial	4	0.10	0.20	3	18	97.12
	RBF	32	0.50	0.10	N/A	10	97.40
PCA-SVR	Linear	1	0.10	N/A	N/A	14	108.77
	Polynomial	4	0.00	0.10	1	20	67.67
	RBF	4	0.00	0.10	N/A	20	67.77
PLSR-SVR	Linear	1	0.10	N/A	N/A	13	111.75
	Polynomial	16	0.00	0.80	1	20	111.90
	RBF	32	0.70	0.30	N/A	3	108.82

4.4.3.5 b) Jalpaiguri

The RBF kernel results in the highest level of 20 support vectors since the ε is zero in all three models. The γ value of greater than zero indicates certain nonlinearity in the model. The same kernel achieved a comparatively lower RMSE among the other two kernels in all three models. There was a reduction in the number of support vectors in polynomial as well as linear kernels. But their RMSE is comparatively high. The PLSR-SVR model with RBF kernel achieved the lowest RMSE of 0.04 among all the kernels of all three models.

4.4.3.5 c) Malda

In SR-SVR model, the RBF kernel with ε of zero, Cost of 32 and γ of 0.10 results in 20 support vectors and the RBF kernel achieved the lowest RMSE among other two kernels of the SR-SVR model. The linear kernel which has ε of 0.10 results in 18 support vectors and comparatively lower RMSE is obtained in linear kernel of the PCA-SVR model. The RBF kernel of PLSR-SVR has the optimum ε of 0.1 and γ of 0.1 which results in 17 support vectors. The RBF kernel achieved comparatively lesser RMSE. The SR-SVR model with RBF kernel achieved the lowest RMSE of 23.10 among all the kernels of all three models.

4.4.3.5 d) Uttar Dinajpur

In SR-SVR model, the polynomial kernel with ε of 0.10, γ of 0.2 and degree of polynomial of 3 results in 18 support vectors and the polynomial kernel achieved the lowest RMSE among other two kernels of the SR-SVR model. Similarly, the polynomial kernel which has ε of zero and γ of 0.10 results in 20 support vectors and comparatively lower RMSE is obtained in polynomial kernel of the PCA-SVR model. The RBF kernel of PLSR-SVR has the optimum ε of 0.70 and γ of 0.30 which results in very low level of 3 support vectors. The RBF kernel achieved comparatively lesser RMSE in PLSR-SVR model. The PCA-SVR model with polynomial kernel achieved the lowest RMSE of 67.67 among all the kernels of all three models.

4.4.4 Penalized Regression Models

The performance of three penalized regression models viz. Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) have been evaluated for each crop in each district. For fitting the penalized regression modes, the penalty parameters Alpha (α) and Lambda (λ) have to be optimized. The alpha value for Ridge and LASSO regression is fixed; that are 0 and 1 respectively. Hence only lambda (λ) value has to be optimized for Ridge and LASSO models. For

ENET model, both alpha (α) and lambda (λ) have to be optimized. The penalty parameters have been optimized using 10-fold cross-validation.

4.4.4.1 Rice

The optimum values of penalty parameters for penalized regression models for rice crop are given for each district in Table 4.44. The lambda values are more in Ridge regression than LASSO in Cooch Behar, Jalpaiguri and Malda. In Uttar Dinajpur, the LASSO penalty is more than Ridge. The better combination of alpha and lambda values of ENET models is optimized. The alpha value for Cooch Behar is close to zero. Hence the ENET model of Cooch Behar is of Ridge type. The alpha value of ENET models of Jalpaiguri, Malda and Uttar Dinajpur are equal to one which is similar to the LASSO model.

Table 4.43 Optimum penalty parameters of penalized regression models for Rice crop

District	Ridge	LASSO	Elastic Net	
	Lambda	Lambda	Alpha	Lambda
Cooch Behar	38	2	0.1	1
Jalpaiguri	39	10	1	10
Malda	33	24	1	24
Uttar Dinajpur	24	49	1	49

The estimated coefficients of three penalized regression models as well as multiple linear regression model with parameters estimated using the Ordinary Least Square (OLS) method are given in Table 4.44. In comparison with OLS estimates, the regression coefficients of each index are shrunken in the Ridge model as it imposes a penalty to the model. But the coefficient values are never shrunken to zero. But in LASSO and ENET models, some of the coefficients are shrunken to zero, thereby variable selection is accomplished. Only important variables are retained in the final LASSO and ENET models.

For Cooch Behar, the sum of absolute coefficient values is lesser in Ridge regression followed by LASSO and ENET. In the LASSO model, the coefficients of three indices namely, WS, PC_Tmax and PC_Tmin are shrunken to zero and the same are removed from the model. In ENET, only PC_RF is shrunken to zero. It can be observed that the estimated coefficients as well as the sum of absolute coefficient values of LASSO and ENET models of Jalpaiguri, Malda and Uttar Dinajpur are equal since the optimum penalty parameters of both models are same. The sum of absolute coefficient values of LASSO and ENET models of Jalpaiguri, Malda and Uttar Dinajpur are lesser than the Ridge model.

Table 4.44 Estimated coefficients of three penalized regression models for Rice crop in four districts

Indices	Cooch Behar				Jalpaiguri				Malda				Uttar Dinajpur			
	OLS	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**
(Intercept)	-27520	-4257	5236	5277	-6654	-6953	134	134	-9018	5854	6038	6038	-1313	554	249	249
Year	1.5	46.9	77.4	78.9	57.6	52.7	66.9	66.9	77.2	46.4	46.9	46.9	58.0	51.2	47.4	47.4
Tmax	-69.1	122.2	5.3	169.3	98.0	170.3	0	0	1734.7	50.9	0	0	153.9	20.2	0	0
Tmin	462.1	17.0	55.0	83.8	16.7	-27.3	0	0	651.6	90.5	0	0	123.0	31.9	0	0
RH	125.4	25.8	-17.9	-42.7	38.4	16.8	0	0	47.5	11.2	0	0	78.3	10.3	0	0
WS	3278.0	413.9	0	48.9	-572.9	-61.7	0	0	205.2	101.1	0	0	242.7	-69.7	0	0
RF	-0.1	-0.1	-0.1	-0.1	-0.1	0.1	0	0	0.3	-0.1	0	0	-0.2	-0.2	0	0
CC_Tmax	-126.3	-17.9	-4.2	-4.4	1.8	-0.3	0	0	-693.4	-81.8	-68.6	-68.6	-16.6	-2.0	0	0
CC_Tmin	-230.0	-6.4	-47.4	-106.0	6.4	0.5	0	0	-317.3	43.4	0	0	37.4	16.5	0	0
CC_RH	62.6	-3.3	-0.6	14.7	-5.2	5.9	3.9	3.9	-256.7	-10.8	0	0	-141.0	-9.3	0	0
CC_WS	500.9	-423.6	-323.1	-137.7	-213.3	-34.8	-8.4	-8.4	-107.0	40.8	33.8	33.8	1.1	-0.2	0	0
CC_RF	-2.2	-0.3	-0.6	-0.8	0.2	0	-0.1	-0.1	3.2	-2.6	-2.8	-2.8	-15.6	-0.2	0	0
PC_Tmax	519.0	28.8	0	-42.7	-20.8	-0.2	0	0	-399.9	-95.3	-26.0	-26.0	-91.9	-6.8	0	0
PC_Tmin	20.3	1.3	0	2.0	-3.3	-1.0	0	0	-598.6	-83.1	0	0	-130.4	20.4	57.5	57.5
PC_RH	-114.0	-3.6	-18.6	-50.0	36.2	21.9	13.1	13.1	151.8	-14.7	-11.7	-11.7	67.7	-7.5	0	0
PC_WS	-832.3	-99.8	-144.1	-316.2	562.0	-81.4	-139.9	-139.9	-7.7	-7.4	-7.8	-7.8	-90.0	-94.0	-1.0	-1.0
PC_RF	6.2	-0.7	-0.1	0	-0.2	-0.1	-0.1	-0.1	-1.7	0.1	0	0	15.3	1.6	0	0
Sum of absolute value of coefficients	**33869**	**5468**	**5930**	**6375**	**8287**	**7428**	**366**	**366**	**14272**	**6534**	**6236**	**6236**	**2575**	**896**	**355**	**355**

Six variables are retained in LASSO and ENET models of Jalpaiguri while seven variables are retained in Malda. For Uttar Dinajpur, only three variables are retained in the final LASSO and ENET models. It can be observed that all the unweighted indices and some of the correlation coefficient-based indices are removed. Path coefficient-based indices are retained in the final model. Due to the coefficient shrinkage as well as variable selection in penalized regression, it is expected that penalized regression models perform better.

4.4.4.2 Wheat

The optimum values of penalty parameters for penalized regression models for wheat crop are given for each district in Table 4.45. The lambda values are more in Ridge regression than LASSO in Cooch Behar, Jalpaiguri and Uttar Dinajpur. In Malda, the LASSO penalty is more than Ridge. The alpha values of ENET models of Cooch Behar and Malda are less than 0.5 which indicates the ENET models of these districts are of Ridge type. Since the alpha value of Uttar Dinajpur is zero and lambda is equal to ridge lambda, the ENET model of Uttar Dinajpur is identical to the Ridge model. Similarly, the alpha value of Jalpaiguri is one and lambda is equal to LASSO lambda, the ENET model of Jalpaiguri is identical to the LASSO model.

Table 4.45 Optimum penalty parameters of penalized regression models for Wheat crop

District	Ridge	LASSO	Elastic Net	
	Lambda	Lambda	Alpha	Lambda
Cooch Behar	400	74	0.3	176
Jalpaiguri	200	58	1	58
Malda	14	60	0.1	4
Uttar Dinajpur	275	38	0	275

The estimated coefficients of three penalized regression models are given in Table 4.46. The sum of absolute coefficient values of all three penalized models is lesser than OLS coefficients in all four districts. It can be observed that the estimated coefficients of the LASSO and ENET models of Jalpaiguri are equal since the optimum penalty parameters of both models are the same. The sum of absolute coefficient values of the LASSO and ENET models are equal and less than the Ridge model. Similarly, the estimated coefficients of the Ridge model and ENET model of Uttar Dinajpur are equal. The sum of absolute coefficient values of the LASSO model is less than the Ridge and ENET models. For Cooch Behar and Uttar Dinajpur districts, the sum of absolute coefficient values is lesser in Ridge regression.

Table 4.46 Estimated coefficients of three penalized regression models for Wheat crop in four districts

Indices	Cooch Behar				Jalpaiguri				Malda				Uttar Dinajpur			
	OLS	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**
(Intercept)	4767	-18	1846	1205	1673	-723	339	339	-1363	68	1725	-308	778	1999	1721	1999
Year	51.83	9.56	13.52	11.24	11.53	18.83	25.38	25.38	42.45	29.91	20.50	39.76	52.16	5.11	5.60	5.11
Tmax	-138.64	21.70	0	0	-10.94	22.95	0	0	123.40	55.0	0	95.94	-21.22	7.40	0	7.40
Tmin	21.87	-4.80	0	0	-10.19	-36.22	0	0	-106.63	-78.25	0	-89.07	66.86	9.15	9.52	9.15
RH	-19.36	11.15	0	4.17	0.20	18.34	0	0	-18.47	-0.97	0	-16.85	20.19	0.73	0	0.73
WS	479.54	167.14	0	0	38.62	239.93	0	0	-273.07	-183.32	0	-303.94	430.42	18.49	0	18.49
RF	-2.74	-0.45	0	0	11.31	-1.36	0	0	-0.85	-0.98	0	-1.32	-2.53	0.33	0	0.33
CC_Tmax	336.96	10.11	0	0	-44.42	-9.30	0	0	15.00	18.29	22.79	7.23	13.33	-6.04	-8.80	6.04
CC_Tmin	-6.48	-31.52	-59.36	-37.08	-90.50	7.35	0	0	12.84	9.59	2.83	10.75	58.60	21.68	46.95	21.68
CC_RH	5.26	9.15	8.34	10.10	13.87	9.41	19.62	19.62	8.11	1.86	0	6.51	-42.15	-0.84	0	0.84
CC_WS	68.03	17.07	0	0	-8.98	-72.40	0	0	125.91	47.64	0	21.82	-285.30	-33.59	0	33.59
CC_RF	2.57	0.02	0	0	-15.36	-7.81	0	0	-4.92	0.06	0	-0.51	-25.84	-5.47	-8.78	5.47
PC_Tmax	-299.91	-9.60	0	0	65.90	2.17	0	0	28.67	25.26	0	26.20	19.77	-12.73	0	12.73
PC_Tmin	-7.50	-8.28	0	-6.26	148.62	0.40	0.91	0.91	-16.82	-16.21	0	-18.75	-1.73	18.67	0	18.67
PC_RH	-4.53	-0.88	0	0	-20.73	9.40	4.05	4.05	12.94	7.23	0	12.69	8.62	-1.50	0	1.50
PC_WS	36.82	35.73	0	19.45	85.35	-95.76	-18.81	-18.81	69.40	58.44	0	149.29	22.84	29.45	17.60	29.45
PC_RF	-13.53	-2.88	0	0	-15.04	0.22	0	0	16.79	9.43	0	14.06	0.21	-1.57	0	1.57
Sum of absolute value of coefficients	**6263**	**358**	**1808**	**1294**	**1850**	**1274**	**408**	**408**	**2239**	**611**	**1771**	**1123**	**2264**	**2172**	**1819**	**2172**

Only three and six variables are retained in the final LASSO and ENET models of Cooch Behar. Whereas, only five variables are retained in the final LASSO and ENET models of Jalpaiguri. Only three variables are retained in the final LASSO model of Malda and none of the indices are shrunken to zero in ENET model. Six variables are retained in the final LASSO and all the variables remain present in the ENET model of Uttar Dinajpur.

4.4.4.3 Potato

The optimum values of penalty parameters for penalized regression models for potato crop are given for each district in Table 4.47. The optimum lambda values of LASSO models of all four districts are one. The lambda values are higher in Ridge regression than LASSO. In Malda, the LASSO penalty is more than Ridge. The alpha values of ENET models of Cooch Behar and Uttar Dinajpur are less than 0.5 which indicates the ENET models of these districts are of Ridge type. Since the alpha value of Malda is one and lambda is equal to LASSO lambda, the ENET model of Malda is identical to the LASSO model.

Table 4.47 Optimum penalty parameters of penalized regression models for Potato crop

District	Ridge	LASSO	Elastic Net	
	Lambda	Lambda	Alpha	Lambda
Cooch Behar	6	1	0.1	4
Jalpaiguri	4	1	0.6	1
Malda	3	1	1	1
Uttar Dinajpur	2	1	0.4	1

The penalized regression models have been fitted using their respective optimum penalty parameter values. The estimated coefficients of three penalized regression models are given in Table 4.48. The sum of absolute coefficient values of all three penalized models is lesser than OLS coefficients in all four districts. Only five variables are retained in the final LASSO model of Cooch Behar, whereas only five variables are removed in the ENET model. The sum of absolute coefficient values is lesser in LASSO followed by ENET and Ridge in Cooch Behar. The sum value is lesser in Ridge followed by ENET and LASSO in Jalpaiguri. Only four and five variables are retained in the final LASSO and ENET models of Jalpaiguri. It can be observed that the estimated coefficients of the LASSO and ENET models of Malda are equal since the optimum penalty parameters of both models are the same. The sum of absolute coefficient values of the LASSO and ENET models are equal and the same is less than the Ridge model.

Table 4.48 Estimated coefficients of three penalized regression models for Potato crop in four districts

Indices	Cooch Behar				Jalpaiguri				Malda				Uttar Dinajpur			
	OLS	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**
(Intercept)	-124.03	19.83	13.86	18.14	-3.94	26.18	50.66	42.58	-224.40	-21.14	-5.78	-5.78	-2.90	-1.65	10.70	8.11
Year	-0.16	0.19	0.18	0.21	0.54	0.25	0.39	0.40	1.34	0.54	0.72	0.72	0.39	0.37	0.49	0.49
Tmax	7.22	0.44	0	0.13	-1.35	0.67	0.03	0.45	-0.04	-0.31	0	0	-0.03	-0.04	0	0
Tmin	9.20	-0.27	0	0	2.11	-0.91	0	0	3.65	-0.93	0	0	0.86	0.43	0	0
RH	1.41	-0.02	0	0	0.29	0.18	0	0	1.06	0.27	0	0	0.91	0.12	0	0
WS	-20.61	-5.30	0	-2.49	26.20	1.53	0	0	12.16	-0.07	0	0	-14.57	0.54	0	0
RF	0.06	0.00	0	0	-0.04	-0.01	0	0	-0.06	-0.01	0	0	-0.04	0.00	0	0
CC_Tmax	-5.38	0.13	0.16	0.12	-0.02	-0.09	0	0	0.47	0.21	0	0	-0.03	-0.17	-0.03	-0.12
CC_Tmin	-3.47	-0.25	-0.47	-0.28	-0.93	-1.14	-2.12	-2.29	-0.50	0.32	0.52	0.52	-0.19	-0.08	-0.13	-0.12
CC_RH	3.40	-0.07	-0.08	-0.07	0.17	-0.01	0	0	-0.01	-0.02	0	0	-1.35	0.05	0	0
CC_WS	6.51	4.01	6.16	3.89	-29.69	0.72	0	0	-54.21	0.56	0	0	54.50	0.68	0	0.23
CC_RF	-0.03	0.00	0	0	-0.01	0.01	0	0	-0.33	0.03	0	0	-1.39	0.21	0.17	0.06
PC_Tmax	-0.72	-0.09	0	-0.07	2.18	0.18	0	0	2.88	0.68	0	0	0.09	-0.32	-0.06	-0.21
PC_Tmin	-2.91	0.13	0	0.09	-3.19	-0.83	0	0	-0.25	-0.06	0	0	-0.03	0.04	0	0.02
PC_RH	-4.01	-0.07	0	-0.06	-0.34	-0.08	-0.05	-0.08	0.11	0.00	0	0	0.44	0.10	0.11	0.12
PC_WS	9.29	2.95	0	2.66	11.56	1.41	0	0.80	43.57	2.18	1.55	1.55	-33.57	0.59	0.30	0.49
PC_RF	0.06	0.00	0	0	0.97	-0.06	0	0	1.64	0.06	0	0	1.63	0.29	0.06	0.27
Sum of absolute value of coefficients	**198.44**	**33.75**	**20.91**	**28.21**	**83.55**	**34.26**	**53.25**	**46.60**	**346.68**	**27.39**	**8.57**	**8.57**	**112.92**	**5.68**	**12.05**	**10.31**

Only time trend variable, CC_Tmin and PC_WS are retained in the final LASSO and ENET models. The sum of absolute coefficient values of the Ridge model is comparatively less followed by ENET model in Uttar Dinajpur. For Cooch Behar and Uttar Dinajpur districts, the sum of absolute coefficient values is lesser in Ridge regression. Eight and ten variables of Uttar Dinajpur are retained in final LASSO and ENET models respectively.

4.4.4.4 Jute

The optimum values of penalty parameters for penalized regression models for jute crop are given for each district in Table 4.49. The optimum lambda values of both Ridge and LASSO models of all four districts are equal to one. The alpha values of ENET models of all four districts are less than 0.5 which indicates the ENET models are of Ridge type. Since the alpha value of Cooch Behar is zero and lambda is equal to ridge lambda, the ENET model of Cooch Behar is identical to the Ridge model.

Table 4.49 Optimum penalty parameters of penalized regression models for Jute crop

District	Ridge	LASSO	Elastic Net	
	Lambda	Lambda	Alpha	Lambda
Cooch Behar	1	1	0	1
Jalpaiguri	1	1	0.1	10
Malda	1	1	0.3	24
Uttar Dinajpur	1	1	0.2	49

The estimated coefficients of three penalized regression models are given in Table 4.50. The sum of absolute coefficient values of all three penalized models are less than OLS coefficients in all four districts. Only time trend variable and CC_WS are retained in final LASSO model of Cooch Behar, whereas RF, CC_RH, CC_RF and PC_RF are removed in the ENET model. The sum of absolute coefficient values is lesser in LASSO in Cooch Behar. Only time trend variable, Tmax, CC_Tmin and PC_RH are retained in the final LASSO model of Jalpaiguri. Along with these three indices, PC_WS also retained in the ENET model. The sum value is lesser in Ridge followed by ENET and LASSO in Jalpaiguri. Only time trend variable and CC_WS are retained in the final LASSO model of Malda. Along with these two indices, PC_WS is also retained in the ENET model. The sum value is lesser in LASSO followed by ENET and Ridge in Malda. Only time trend variable and CC_WS are retained in the final LASSO model of Uttar Dinajpur. Seven variables are retained in the final ENET model of Uttar Dinajpur. The sum of absolute coefficient values of Ridge model is comparatively less followed by ENET model in Uttar Dinajpur.

Table 4.50 Estimated coefficients of three penalized regression models for Jute crop in four districts

Indices	Cooch Behar				Jalpaiguri				Malda				Uttar Dinajpur			
	OLS	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**
(Intercept)	-1.38	13.99	10.59	13.99	-3.94	26.18	50.66	42.58	-68.01	15.1	12.6	14.31	41.01	3.59	10.59	7.09
Year	0.14	0.1	0.12	0.1	0.54	0.25	0.39	0.40	0.20	0.16	0.14	0.16	0.21	0.15	0.12	0.16
Tmax	-1.18	0.05	0	0.05	-1.35	0.67	0.03	0.45	1.71	-0.02	0	0	-0.28	0.05	0	0
Tmin	7.25	-0.1	0	-0.1	2.11	-0.91	0	0	-0.02	0.07	0	0	-1.11	0.04	0	0.01
RH	0.38	0.03	0	0.03	0.29	0.18	0	0	0.50	0.02	0	0	0.02	0.02	0	0
WS	-1.75	-2.49	0	-2.49	26.20	1.53	0	0	2.53	0.46	0	0	-3.32	-0.48	0	0
RF	0.00	0	0	0	-0.04	-0.01	0	0	-0.01	0	0	0	0.00	0	0	0
CC_Tmax	1.02	-0.11	0	-0.11	-0.02	-0.09	0	0	-0.12	-0.06	0	0	-0.59	0.04	0	0.01
CC_Tmin	-5.99	-0.11	0	-0.11	-0.93	-1.14	-2.12	-2.29	0.08	0.02	0	0	1.06	0.03	0	0.03
CC_RH	0.07	0	0	0	0.17	-0.01	0	0	0.07	-0.01	0	0	-0.09	0	0	0
CC_WS	-2.20	0.26	-0.05	0.26	-29.69	0.72	0	0	-3.09	-0.42	-0.02	-0.32	0.41	-0.17	-0.05	-0.17
CC_RF	0.02	0	0	0	-0.01	0.01	0	0	0.02	-0.01	0	0	0.04	0	0	0
PC_Tmax	1.02	0.18	0	0.18	2.18	0.18	0	0	-0.17	-0.02	0	0	0.58	0.1	0	0.09
PC_Tmin	-2.25	-0.08	0	-0.08	-3.19	-0.83	0	0	-0.14	-0.01	0	0	-0.17	0.01	0	0
PC_RH	-0.34	0.02	0	0.02	-0.34	-0.08	-0.05	-0.08	-0.08	-0.01	0	0	0.04	0	0	0
PC_WS	3.33	0.16	0	0.16	11.56	1.41	0	0.80	1.08	-0.49	0	-0.23	-1.47	-0.3	0	-0.24
PC_RF	0.00	0	0	0	0.97	-0.06	0	0	0.00	0	0	0	-0.10	0	0	0
Sum of absolute value of coefficients	28.35	17.68	10.76	17.68	83.55	34.26	53.25	46.60	77.84	16.88	12.76	15.02	50.49	4.98	10.76	7.80

4.4.4.5 Rapeseed-mustard

The optimum values of penalty parameters for penalized regression models for jute crop are given for each district in Table 4.51. The lambda values are more in Ridge regression than LASSO. The alpha values for Cooch Behar and Jalpaiguri districts are less than 0.5. Hence the ENET models for these districts are of Ridge type. The ENET models for Malda and Uttar Dinajpur are of LASSO type as their alpha values are near one.

Table 4.51 Optimum penalty parameters of penalized regression models for Rapeseed-Mustard crop

District	Ridge	LASSO	Elastic Net	
	Lambda	Lambda	Alpha	Lambda
Cooch Behar	299	25	0.3	66
Jalpaiguri	70	12	0.3	20
Malda	62	12	0.8	14
Uttar Dinajpur	130	27	0.6	38

The estimated coefficients of three penalized regression models are given in Table 4.52. The sum of absolute coefficient values of all three penalized models is lesser than OLS coefficients in all four districts. In LASSO and ENET models of Cooch Behar, only four and six indices are retained. The sum value is comparatively lower in Ridge regression followed by LASSO and ENET. For Jalpaiguri, at the optimum penalty values of LASSO and ENET models, eight and twelve indices are retained. The sum of absolute coefficient value is comparatively lower in ENET followed by LASSO and Ridge model. Only six indices are retained at the optimum lambda level of LASSO and ENET models of Malda and Uttar Dinajpur districts. The sum of absolute coefficient value is comparatively lower in LASSO followed by ENET and Ridge model.

Table 4.52 Estimated coefficients of three penalized regression models for Rapeseed-mustard crop in four districts

Indices	Cooch Behar				Jalpaiguri				Malda				Uttar Dinajpur			
	OLS	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**	**OLS**	**Ridge**	**LASSO**	**ENET**
(Intercept)	2710	22	-21	-287	1634	486	438	40	1712	1004	868	896	12944	-1846	-1000	-1187
Year	19.31	0.50	0	0	20.38	6.53	10.28	9.23	24.24	11.66	15.57	17.21	134.99	6.84	16.53	15.12
Tmax	22.79	3.12	0	0	-213.91	13.29	0	10.56	19.96	-3.19	0	0	291.31	14.5	0	0
Tmin	-217.13	-2.20	0	0	-58.14	-33.21	-10.63	-29.99	-31.14	-1.98	0	0	-65.99	5.38	0	0
RH	-71.42	-0.97	0	0	-56.74	2.68	0	0	-0.03	1.54	0	0	-71.18	3.61	1.41	5.17
WS	-415.22	-21.58	0	0	-505.39	25.88	0	0	-151.76	16.81	0	0	927.95	94.01	0	0
RF	-2.31	-0.01	0	0	0.59	-0.39	0	0	2.69	-0.39	0	0	-23.56	-0.82	0	0
CC_Tmax	75.41	5.00	8.56	6.46	30.58	5.57	0	3.63	-19.00	3.4	0	0	-118.28	15.88	18.57	23.71
CC_Tmin	706.91	-5.10	0	0	-27.17	-4.65	-0.11	-5.55	3.88	-1.68	0	0	-56.49	13.76	2.16	3.91
CC_RH	-11.03	1.03	4.42	1.93	30.37	-1.56	0	-0.15	0.10	0.12	0.14	0.14	-13.56	6.66	9.63	10.27
CC_WS	1811.70	61.14	137.69	99.79	589.54	-40.8	0	-0.98	-329.36	-48.35	-47.12	-47.87	606.34	-19.71	0	0
CC_RF	-10.08	0.14	0	0	0.81	-0.56	-0.73	-0.69	12.54	1.61	2.20	2.21	-153.46	-2.63	0	0
PC_Tmax	-6.54	9.03	13.3	12.94	245.19	10.78	26.54	21.06	29.04	5.68	0	0	-259.54	-2.83	0	0
PC_Tmin	-529.94	-5.50	0	-0.04	-19.43	-14.19	-33.26	-16.51	-80.51	-27	-10.83	-12.45	-241.14	21.48	13.49	19.57
PC_RH	35.20	1.22	0	0	5.28	0.56	0.52	0.62	5.87	3.73	3.73	3.81	4.21	2.87	0	0
PC_WS	-1376.56	79.88	0	74.34	-23.32	-26.64	-19.24	-23.38	215.18	-14.71	0	0	-1300.38	17.15	0	0
PC_RF	41.96	0.71	0	0	8.68	-0.99	0	0	-24.95	1.51	0	0	16.33	1.12	0	0
Sum of absolute value of coefficients	**8063**	**219**	**380**	**482**	**3469**	**674**	**539**	**532**	**2662**	**1147**	**948**	**980**	**17229**	**2076**	**1062**	**1265**

4.5 Comparison of Performance of Fitted Models

The MLR, ANN and SVR models have been fitted under three variable selection or dimension reduction techniques *viz.* Stepwise Regression (SR), Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR). Along with these, three penalized regression models are also fitted. Hence, twelve models have been fitted for each crop in each district. Among the fitted models, the best-fitted model for each crop in each district has been selected by examining the following model evaluation criteria such as R^2, MAE, RMSE and nRMSE.

The data from the years 1997-98 to 2019-20 have been utilized. Among that 80% of the data between the years 1997-98 and 2019-20 are randomly selected for model training (calibration) and the remaining 20% of data are utilized for validation of the fitted models using residual measures such as MAE, RMSE and nRMSE. The model which performs better in both the training and validation stages is considered the best-fitted model. The best-fitted model for each crop in each district is selected.

4.5.1 Rice

The model evaluation criteria of fitted models for rice crop for both calibration as well as validation stages are given in Table 4.53. It can be observed that MLR, ANN and SVR models under all three variable selection methodologies have having highest R^2 of close to one. However, error measures such as MAE, RMSE and nRMSE are very high for both calibration as well as validation stages for all the districts. But penalized regression models are having moderate R^2. But the MAE, RMSE and nRMSE are comparatively low for both calibration as well as validation stages.

For Cooch Behar, the SR-SVR model has highest R^2 of 0.99 and low MAE, RMSE and nRMSE of 4.32, 18.85 and 0.01 respectively. But the error measures are very high for validation data. It is clearly indicating the presence of overfitting issue. The overfitted model describe the training data very well; but its performance for new data is generally poor. The Ridge model which has R^2 of 0.72 is having lowest MAE, RMSE and nRMSE of 2.69, 3.46 and 0.14 for training data respectively. The MAE, RMSE and nRMSE for testing data are also lowest level of 0.55, 6.02 and 0.30. From the Fig. 4.16, it can also be seen that RMSE of Ridge model during calibration as well as validation stages are low. Other two penalized regression models *viz.* ENET and LASSO models are also having moderate R^2 and less error in both training and validation stages. Hence, the Ridge model is found to perform well for predicting the yield of rice crop in Cooch Behar followed by ENET and LASSO models.

Table 4.53 The model evaluation parameters in calibration and validation stages of fitted models for Rice crop

Model	Calibration				Validation			Calibration				Validation		
	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE
	Cooch Behar							**Jalpaiguri**						
SR-MLR	0.99	41.9	57.53	0.03	24.97	26.63	0.01	0.99	38.03	48.96	0.02	42.07	42.99	0.03
SR-ANN	0.98	50.85	68.51	0.04	410.63	478.96	0.37	0.99	39.37	50.14	0.03	484.82	611.33	0.04
SR-SVR	0.99	4.32	18.85	0.01	243.86	247.13	0.14	0.99	42.33	43.73	0.02	229.07	273.79	0.16
PCA-MLR	0.83	156.34	194.43	0.10	168.46	205.55	0.11	0.80	170.04	206.74	0.10	309.65	353.74	0.21
PCA-ANN	0.85	147.48	181.13	0.11	53.96	64.03	0.05	0.83	154.99	192.28	0.13	309.52	328.75	0.25
PCA-SVR	0.82	140.59	206.09	0.10	321.40	331.73	0.18	0.81	189.79	202.53	0.10	308.08	319.79	0.19
PLSR-MLR	0.98	43.38	60.34	0.03	74.54	75.12	0.04	0.99	45.22	56.42	0.03	40.97	48.9	0.03
PLSR-ANN	0.98	50.64	65.84	0.16	867.32	1212.32	0.46	0.98	51.99	60.25	0.04	269.47	290.36	0.22
PLSR-SVR	0.98	39.32	63.05	0.03	58.03	61.31	0.03	0.98	45.53	58.61	0.03	47.81	59.94	0.04
Ridge	0.72	2.69	3.46	0.14	0.55	6.02	0.30	0.86	1.55	2.10	0.08	0.88	3.33	0.16
LASSO	0.67	2.96	3.73	0.15	0.52	6.29	0.31	0.83	1.87	2.29	0.09	0.87	3.46	0.17
ENET	0.70	2.79	3.57	0.15	0.57	6.11	0.30	0.88	1.51	1.94	0.08	0.90	3.30	0.16
	Malda							**Uttar Dinajpur**						
SR-MLR	0.92	93.91	121.71	0.05	73.91	86.16	0.03	0.94	91.55	103.63	0.05	114.88	123.31	0.06
SR-ANN	0.92	93.01	124.52	0.08	361.82	364.55	0.23	0.94	93.6	104.52	0.07	338.88	385.78	0.29
SR-SVR	0.94	41.09	111.01	0.04	290.61	314.7	0.13	0.97	74.07	76.54	0.03	59.6	62.66	0.03
PCA-MLR	0.76	159.98	212.12	0.08	34.14	42.29	0.01	0.70	168.16	232.29	0.11	600.25	604.20	0.29
PCA-ANN	0.89	115.05	143.75	0.09	521.33	599.38	0.37	0.83	112.77	172.13	0.11	712.01	807.78	0.61
PCA-SVR	0.75	154.65	216.84	0.08	41.68	54.04	0.02	0.83	121.69	175.21	0.08	308.77	318.20	0.15
PLSR-MLR	0.93	78.17	109.41	0.04	62.87	63.59	0.03	0.96	72.35	83.62	0.04	73.81	92.41	0.04
PLSR-ANN	0.94	78.63	106.44	0.07	188.62	216.05	0.13	0.96	7.00	2.34	0.05	207.78	256.69	0.19
PLSR-SVR	0.93	69.07	116.61	0.04	84.85	95.9	0.04	0.96	74.51	86.35	0.04	56.99	63.25	0.03
Ridge	0.85	2.82	3.58	0.13	0.59	7.89	0.33	0.89	1.61	1.90	0.08	0.90	2.25	0.10
LASSO	0.82	2.93	3.96	0.15	0.59	8.01	0.34	0.82	2.02	2.44	0.10	0.82	2.47	0.11
ENET	0.82	2.93	3.96	0.15	0.59	8.01	0.34	0.90	1.54	1.85	0.08	0.83	2.16	0.09

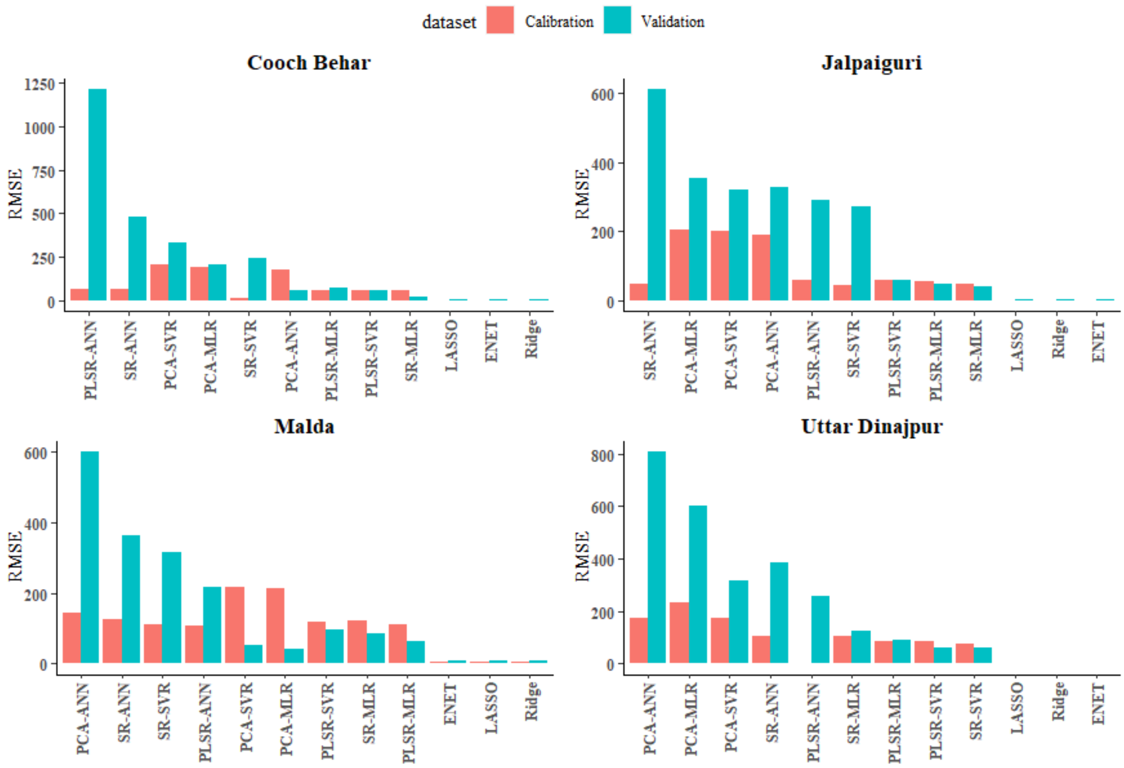

Fig. 4.16 RMSE plots of all fitted models of Rice crop of four districts

Similarly, for Jalpaiguri, the MLR, ANN and SVR models are having high R^2. But the error measures are very high for both training and validation data. The ENET model which has R^2 of 0.88 is having lowest MAE, RMSE and nRMSE of 1.51, 1.94 and 0.08 for training data respectively. The MAE, RMSE and nRMSE for testing data are also lowest level of 0.90, 3.30 and 0.16. From Fig.4.16, it can also be seen that RMSE of ENET model during calibration as well as validation stages are low. Other two penalized regression models *viz.* Ridge and LASSO models also have moderate R^2 and less error in both the training and validation stages. Hence the ENET model is found to perform well for predicting the yield of rice crop in Jalpaiguri followed by the Ridge and LASSO models.

The MLR, ANN and SVR models have high R^2 in Malda. But the error measures are very high for both training and validation data. The Ridge model which has R^2 of 0.85 is having lowest MAE, RMSE and nRMSE of 2.82, 3.58 and 0.13 for training data respectively. The MAE, RMSE and nRMSE for testing data are also the lowest level of 0.59, 7.89 and 0.33. From Fig.4.16**,** it can also be seen that RMSE of the Ridge model during calibration as well as validation stages are low. Other two penalized regression models *viz.* ENET and LASSO also have moderate R^2 and less error in both the training and validation stages. Hence the Ridge model is found to perform well for predicting the yield of rice crop in Malda followed by the ENET and LASSO models.

For Uttar Dinajpur, the PLSR-ANN model has the highest R^2 of 0.96 and low MAE, RMSE and nRMSE of 7.00, 3.14 and 0.05 respectively. But the error measures are very high for validation data due the possibility of overfitting issue. The ENET model which has R^2 of 0.90 is having lowest MAE, RMSE and nRMSE of 1.54, 1.85 and 0.08 for training data respectively. The MAE, RMSE and nRMSE for testing data are also lowest level of 0.83, 2.16 and 0.09. From the Fig.4.16**,** it can also be seen that the RMSE of ENET model during calibration as well as validation stages are low. Other two penalized regression models *viz.* Ridge and LASSO models are also have moderate R^2 and less error in both the training and validation stages. Hence the ENET model is found to perform well for predicting the yield of rice crop in Uttar Dinajpur Ridge and LASSO models.

The penalized regression models were showing stable performance in both calibration and validation of the model (Sridhara *et al* 2023). The penalized regression models that accompanied by coefficient shrinkage and variable selection are found to be the best-fitted models (Ajith *et al* 2023).

4.5.2 Wheat

The model evaluation criteria of fitted models for wheat crop for both calibration as well as validation stages are given in Table 4.54. For Cooch Behar, the PCA-ANN and PLSR-ANN models have the highest R^2 of 0.90. The MAE, RMSE and nRMSE of PCA-ANN are 86.82, 110.25 and 0.07 respectively. But the error measures are very high for testing data. The PLSR-ANN model has MAE, RMSE and nRMSE of 92.82, 113.60 and 0.08 for training data and 76.05, 87.30 and 0.20 for testing data respectively. From Fig. 4.17, it can also be seen that RMSE of PLSR-ANN model during calibration as well as validation stages are low. The PCA-ANN model is also having R^2 of 0.90 and less error in both the training and validation stages. Hence, the PLSR-ANN model is found to perform well for predicting the yield of wheat crop in Cooch Behar followed by PCA-ANN. The ANN algorithm using PLSR dimension reduction simplifies the neural network structure and increases the predictive efficiency (Jia *et al* 2014). The best fitted PLSR-ANN model using Tangent hyperbolic (Tanh) function as the activation function and PLSR components as the input is given in Fig. 4.18.

Table 4.54 The model evaluation parameters in calibration and validation stages of fitted models for Wheat crop

Model	Calibration				Validation			Calibration				Validation		
	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE
Cooch Behar								**Jalpaiguri**						
SR-MLR	0.74	136.01	180.40	0.09	129.13	148.78	0.09	0.89	116.46	148.40	0.07	119.68	130.60	0.08
SR-ANN	0.76	139.30	175.60	0.12	86.64	96.36	0.55	0.95	77.14	105.44	0.05	95.08	100.79	0.50
SR-SVR	0.75	114.52	178.17	0.09	142.01	187.42	0.12	0.96	40.21	92.18	0.04	61.94	62.24	0.16
PCA-MLR	0.67	162.83	204.84	0.09	221.54	322.9	0.20	0.84	142.19	179.92	0.08	130.07	157.75	0.09
PCA-ANN	0.90	86.82	110.25	0.07	94.45	95.08	0.54	0.94	78.22	113.27	0.05	96.68	97.55	0.49
PCA-SVR	0.96	35.65	77.43	0.04	211.91	217.05	0.14	0.83	138.17	191.56	0.09	130.14	157.68	0.10
PLSR-MLR	0.83	118.71	146.62	0.07	224.00	248.86	0.15	0.91	102.41	136.99	0.06	112.35	150.27	0.09
PLSR-ANN	0.90	92.82	113.60	0.08	76.05	87.30	0.20	0.91	109.70	138.11	0.07	145.92	149.53	0.75
PLSR-SVR	0.81	111.81	156.75	0.08	225.12	239.54	0.15	0.91	95.26	140.27	0.06	121.84	151.79	0.09
Ridge	0.74	139.01	182.70	0.08	232.67	294.15	0.18	0.89	115.09	146.31	0.06	101.43	121.62	0.07
LASSO	0.65	160.97	208.53	0.10	256.68	328.42	0.20	0.85	137.43	174.12	0.07	272.12	313.75	0.19
ENET	0.66	158.47	205.81	0.09	239.39	306.37	0.19	0.85	137.43	174.12	0.07	272.12	313.75	0.19
Malda								**Uttar Dinajpur**						
SR-MLR	0.81	97.75	120.62	0.04	67.96	76.65	0.03	0.81	95.92	109.73	0.05	121.56	146.29	0.06
SR-ANN	0.95	45.83	60.21	0.06	225.77	228.18	0.26	0.91	58.52	74.36	0.08	156.32	179.36	0.35
SR-SVR	0.95	34.26	63.12	0.02	125.37	144.26	0.06	0.86	90.60	95.34	0.04	124.86	164.56	0.07
PCA-MLR	0.73	118.36	144.40	0.05	152.59	161.62	0.06	0.81	80.81	111.50	0.04	104.82	134.24	0.05
PCA-ANN	0.95	45.63	60.14	0.06	207.65	226.69	0.26	0.93	51.45	68.69	0.08	178.20	220.95	0.43
PCA-SVR	0.99	3.56	15.64	0.01	235.63	291.71	0.11	0.79	79.31	116.51	0.05	100.55	140.13	0.06
PLSR-MLR	0.92	62.98	79.58	0.03	97.61	102.17	0.04	0.88	70.38	89.20	0.04	99.49	121.26	0.05
PLSR-ANN	0.96	42.30	52.04	0.05	330.44	395.27	0.45	0.99	17.99	30.09	0.03	254.36	312.97	0.61
PLSR-SVR	0.91	60.80	82.23	0.03	81.53	84.17	0.03	0.96	22.95	52.95	0.02	90.61	120.18	0.05
Ridge	0.92	62.01	73.86	0.02	102.46	106.25	0.04	0.73	92.22	121.71	0.05	141.92	163.96	0.07
LASSO	0.63	132.69	167.81	0.06	242.33	265.81	0.10	0.70	95.86	128.54	0.05	144.30	157.11	0.07
ENET	0.96	39.63	48.79	0.01	122.80	146.50	0.05	0.73	92.22	121.71	0.05	141.92	163.96	0.07

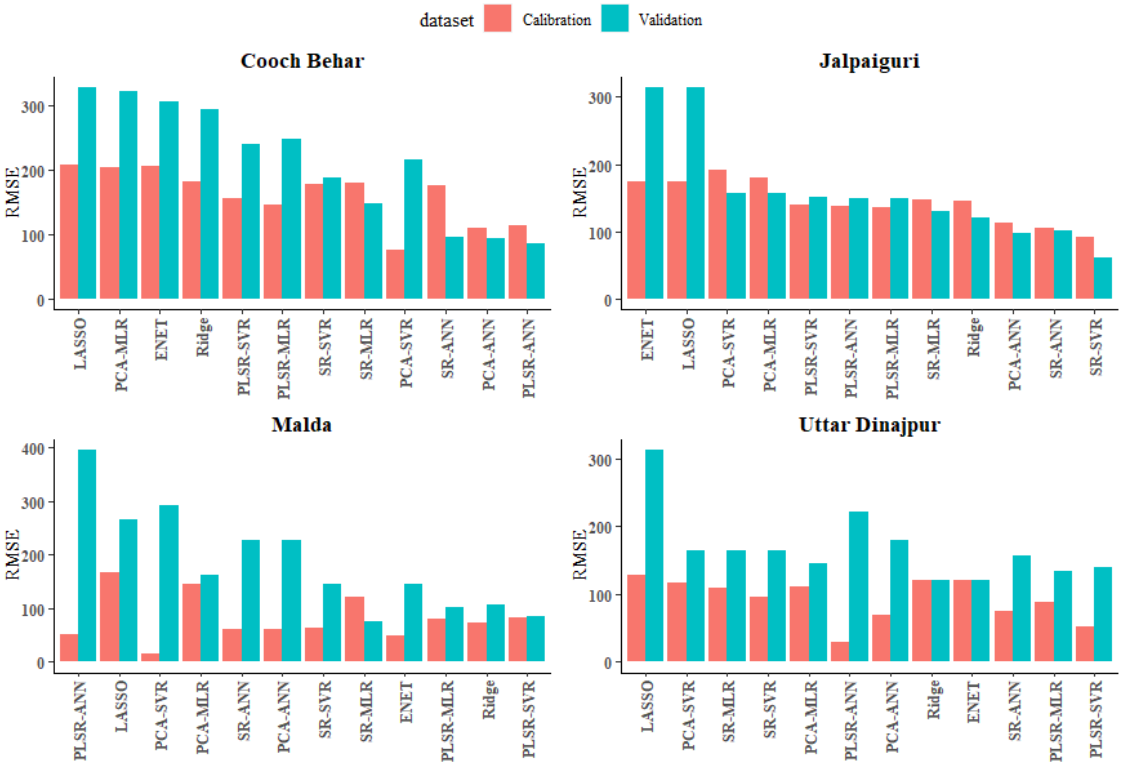

Fig. 4.17 RMSE plots of all fitted models of Wheat crop of four districts

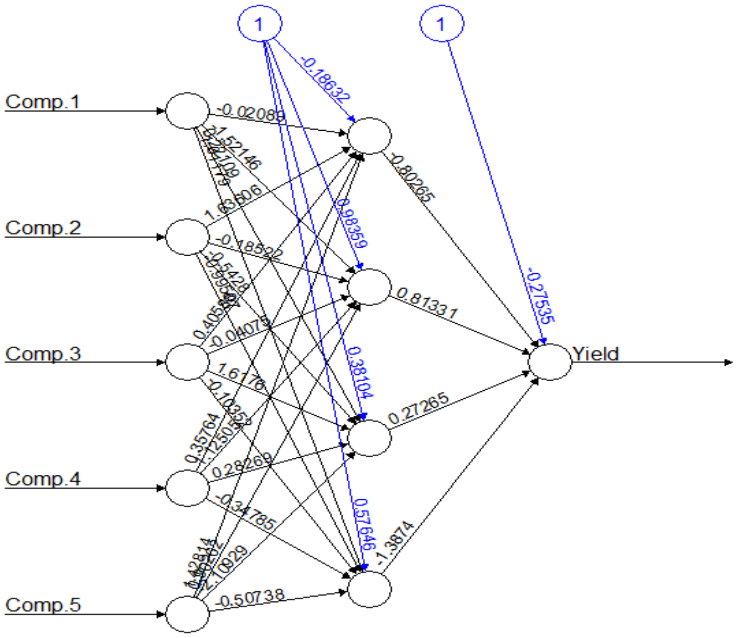

Fig. 4.18 The best fitted PLSR-ANN model for Wheat crop of Cooch Behar district

For Jalpaiguri, the SR-SVR model has the highest R^2 of 0.96. The error measures such as MAE, RMSE and nRMSE are comparatively low level of 40.21, 92.18 and 0.04 for training data and 61.94, 62.24 and 0.16 testing data respectively. From the Fig. 4.17, it can also be seen that RMSE of SR-SVR model during the calibration as well as validation stages are low. The SR-ANN and PCA-ANN models are also have high R^2 and low less in both the training and validation stages. Hence, the SR-SVR model is found to perform well in predicting the yield of wheat crop in Jalpaiguri followed by SR-ANN and PCA-ANN. Due to the presence of a nonlinear complex association between crop yield and weather factors, the non-linear machine linear models accompanied by feature selection is effective for crop yield prediction (Shahsavari *et al* 2023).

For Malda, PCA-SVR, PLSR-ANN, PCA-ANN, SR-ANN and SR-SVR models are having high R^2. But the error measures are very high for both training and validation data. The PLSR-SVR model which has R^2 of 0.91 is having lowest MAE, RMSE and nRMSE of 60.80, 82.23 and 0.03 for training data respectively. The MAE, RMSE and nRMSE for testing data are also lowest level of 81.53, 84.17 and 0.03. From the Fig. 4.17, it can also be seen that RMSE is low in PLSR-SVR model during both the calibration as well as validation stages. The Ridge model and PLSR-MLR models also have high R^2 and less error in both the training and validation stages. Hence, the PLSR-SVR model is found to perform well in predicting the yield of wheat crop in Malda followed by Ridge regression and PLSR-MLR models. The hyperparameter-tuned Support Vector model performs better for predicting crop yield (Saha *et al* 2021).

For Uttar Dinajpur, the PLSR-ANN model has the highest R^2 of 0.99. The MAE, RMSE and nRMSE of PLSR-ANN are 17.99, 30.09 and 0.03 respectively. However, the error measures are very high for testing data. The PLSR-SVR model which has R^2 of 0.96 has MAE, RMSE and nRMSE of 22.95, 52.95 and 0.02 for training data and 90.61, 120.18 and 0.05 for testing data respectively. From Fig. 4.17, it can also be seen that RMSE of PLSR-SVR model during calibration as well as validation stages are low. The PLSR-MLR model also has R^2 of 0.88 and less error in both the training and validation stages. Hence the PLSR-SVR model is found to perform well for predicting the yield of wheat crop in Uttar Dinajpur followed by PLSR-MLR.

4.5.3 Potato

The model evaluation criteria of fitted models for potato crop for both calibration as well as validation stages are given in Table 4.55. For Cooch Behar, SR-SVR, PCA-ANN and SR-ANN models are having high R^2. The MAE, RMSE and nRMSE are also low for training data. But the error measures are very low for testing data which indicates the chance of overfitting issue. The PLSR-ANN model which has R^2 of 0.79 is having lowest MAE, RMSE and nRMSE of 1.98, 1.64 and 0.13 for training data respectively. The MAE, RMSE and nRMSE for testing data are also lowest level of 1.56, 1.79 and 0.10. From Fig. 4.19, it can also be seen that RMSE of PLSR-ANN model during calibration as well as validation stages are low. The PCA-MLR and penalized regression models also have high R^2 and less error in both the training and validation stages. Hence, the PLSR-ANN model is found to perform well in predicting the yield of the potato crop of Cooch Behar followed by PCA-MLR and penalized regression models. The ANN model using PLSR components as input performs better in a prediction context (Sharabiani *et al* 2023).

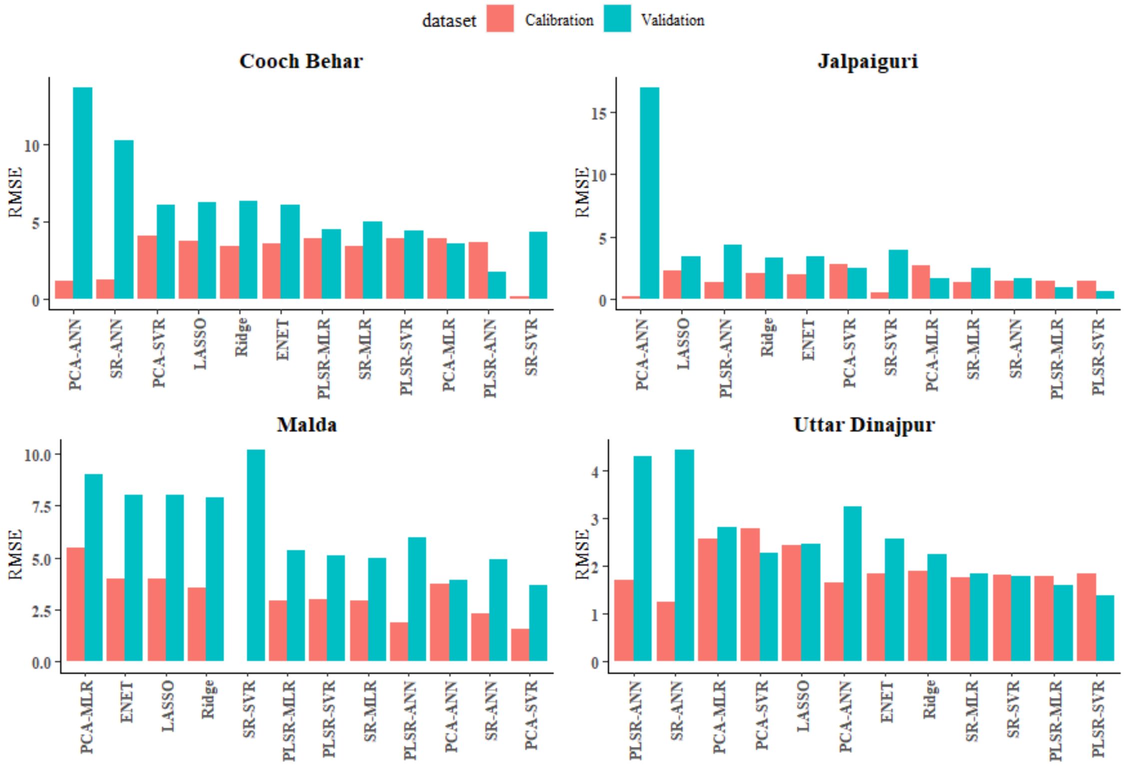

Fig. 4.19 RMSE plots of all fitted models of Potato crop of four districts

For Jalpaiguri, the PCA-ANN model has the highest R^2 of 0.99. The MAE, RMSE and nRMSE are 0.13, 0.24 and 0.01 respectively. But the error measures are very high for testing data. The PLSR-SVR model which has R^2 of 0.92 has MAE, RMSE and nRMSE of 1.14, 1.44 and 0.06 for training data and 0.52, 0.61 and 0.03 for testing data respectively. From Fig. 4.19, it can also be seen that RMSE of PLSR-SVR model during calibration as well as validation stages are low. The PLSR-MLR, SR-ANN and SR-MLR models are also have high R^2 and less error in both the training and validation stages. Hence, the PLSR-SVR model is found to perform well for predicting the yield of potato crop in Jalpaiguri followed by PLSR-MLR, SR-ANN and SR-MLR models.

For Malda, the SR-SVR model has the highest R^2 of 0.99. The MAE, RMSE and nRMSE of zero. But the error measures are very high for testing data. The PCA-SVR model which has R^2 of 0.97 has MAE, RMSE and nRMSE of 0.35, 1.56 and 0.06 for training data and 3.36, 3.65 and 0.15 for testing data respectively.

Table 4.55 The model evaluation parameters in calibration and validation stages of fitted models for Potato crop

Model	Calibration				Validation			Calibration				Validation		
	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE
	Cooch Behar							**Jalpaiguri**						
SR-MLR	0.73	2.63	3.40	0.14	3.13	4.99	0.24	0.94	1.13	1.38	0.05	2.46	2.52	0.12
SR-ANN	0.96	1.14	1.26	0.06	8.95	10.27	0.55	0.93	1.20	1.45	0.07	0.53	1.63	0.04
SR-SVR	0.99	0.05	0.22	0.01	4.14	4.33	0.21	0.99	0.20	0.53	0.02	3.63	4.00	0.19
PCA-MLR	0.64	2.92	3.89	0.15	2.59	3.61	0.17	0.76	2.18	2.71	0.11	1.62	1.66	0.08
PCA-ANN	0.97	0.94	1.19	0.05	11.28	13.66	0.73	0.99	0.13	0.24	0.01	14.21	16.97	0.99
PCA-SVR	0.62	2.80	4.08	0.17	4.44	6.05	0.30	0.75	2.15	2.77	0.11	2.55	2.55	0.12
PLSR-MLR	0.64	2.85	3.90	0.16	3.44	4.55	0.22	0.93	1.32	1.52	0.06	0.91	0.95	0.05
PLSR-ANN	0.79	1.98	1.64	0.13	1.56	1.79	0.10	0.94	1.09	1.34	0.07	4.28	4.36	0.25
PLSR-SVR	0.64	2.83	3.91	0.16	3.36	4.41	0.22	0.92	1.14	1.44	0.06	0.52	0.61	0.03
Ridge	0.72	2.69	3.46	0.14	0.59	6.32	0.31	0.86	1.55	2.10	0.08	0.88	3.33	0.16
LASSO	0.67	2.96	3.73	0.15	0.52	6.29	0.31	0.83	1.87	2.29	0.09	0.87	3.46	0.17
ENET	0.70	2.79	3.57	0.15	0.57	6.11	0.30	0.88	1.51	1.94	0.08	0.90	3.40	0.16
	Malda							**Uttar Dinajpur**						
SR-MLR	0.90	2.46	2.94	0.11	4.31	4.99	0.21	0.91	1.42	1.76	0.07	1.53	1.85	0.08
SR-ANN	0.94	1.83	2.31	0.08	4.56	4.92	0.21	0.95	0.98	1.25	0.07	4.13	4.42	0.32
SR-SVR	0.99	0.00	0.00	0.00	7.25	10.16	0.43	0.90	1.43	1.81	0.08	1.34	1.78	0.08
PCA-MLR	0.65	4.14	5.49	0.20	7.06	8.97	0.38	0.80	2.22	2.56	0.11	2.30	2.82	0.13
PCA-ANN	0.84	2.66	3.74	0.13	3.02	3.92	0.17	0.92	1.20	1.64	0.09	2.69	3.25	0.23
PCA-SVR	0.97	0.35	1.56	0.06	3.36	3.65	0.15	0.76	2.62	2.79	0.12	1.94	2.28	0.10
PLSR-MLR	0.90	2.50	2.91	0.11	4.59	5.32	0.22	0.90	1.56	1.79	0.07	1.37	1.60	0.07
PLSR-ANN	0.96	1.44	1.88	0.07	5.52	5.98	0.25	0.91	1.36	1.71	0.10	3.00	4.30	0.31
PLSR-SVR	0.90	2.51	2.99	0.11	4.23	5.10	0.22	0.90	1.52	1.85	0.08	1.13	1.39	0.06
Ridge	0.85	2.82	3.58	0.13	0.59	7.89	0.33	0.89	1.61	1.90	0.08	0.90	2.25	0.10
LASSO	0.82	2.93	3.96	0.15	0.59	8.01	0.34	0.82	2.02	2.44	0.10	0.82	2.47	0.11
ENET	0.82	2.93	3.96	0.15	0.59	8.01	0.34	0.90	1.54	1.85	0.08	0.83	2.56	0.12

From Fig. 4.19, it can also be seen that RMSE are low in PCA-SVR model during calibration as well as validation stages. The R^2 of SR-ANN, PCA-ANN, PLSR-ANN and SR-MLR models are also high and the errors are less in both training and validation stages. Hence, the PCA-SVR model is found to perform well for predicting the yield of potato crop in Malda followed by SR-ANN, PCA-ANN, PLSR-ANN and SR-MLR models

For Uttar Dinajpur, the PLSR-SVR, PLSR-MLR, SR-SVR and SR-SVR models have the R^2 of around 0.90. The MAE, RMSE and nRMSE of both calibration and training stages of the models are very close. But the error measures are very high for testing data. The PLSR-SVR model which has R^2 of 0.90 has comparatively low MAE, RMSE and nRMSE of 1.13, 1.39 and 0.06 for testing data respectively. From Fig. 4.19, it can also be seen that RMSE of PLSR-SVR model during calibration as well as validation stages are less. Hence the PLSR-SVR model is found to perform well for predicting the yield of potato crop in Uttar Dinajpur followed by PLSR-MLR, SR-SVR and SR-SVR models

4.5.4 Jute

The model evaluation criteria of fitted models for jute crop for both calibration as well as validation stages are given in Table 4.56. For Cooch Behar, The PLSR-SVR model which has R^2 of 0.97 is having low MAE, RMSE and nRMSE of 0.19, 0.25 and 0.03 for training data and 0.63, 0.34 and 0.06 for testing data respectively. From Fig. 4.20, it can also be seen that RMSE of PLSR-SVR model during calibration as well as validation stages are low. The SR-SVR, SR-ANN, PLSR-ANN and SR-MLR models also have high R^2 and less error in both training and validation stages. Hence, the PLSR-SVR model is found to perform well for predicting the yield of jute crop in Cooch Behar followed by SR-SVR, SR-ANN, PLSR-ANN and SR-MLR models.

For Jalpaiguri, the SR-SVR model has the highest R^2 of 0.99. The MAE, RMSE and nRMSE are low level of 0.06, 0.12 and 0.01 for training data and 0.44, 0.52 and 0.08 for testing data respectively. From the Fig. 4.20, it can also be seen that RMSE of SR-SVR model during calibration as well as validation stages are low. The PLSR-SVR, PLSR-MLR and SR-MLR models are also have high R^2 and less error in both the training and validation stages. Hence, the SR-SVR model is found to perform well for predicting the yield of jute crop in Jalpaiguri followed by PLSR-SVR, PLSR-MLR and SR-MLR models.

Table 4.56 The model evaluation parameters in calibration and validation stages of fitted models for Jute crop

Model	Calibration				Validation			Calibration				Validation		
	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE
	Cooch Behar							**Jalpaiguri**						
SR-MLR	0.95	0.27	0.35	0.03	0.43	0.45	0.04	0.96	0.33	0.39	0.03	0.52	0.57	0.05
SR-ANN	0.96	0.25	0.32	0.06	0.66	0.76	0.15	0.96	0.37	0.40	0.05	1.18	1.25	0.32
SR-SVR	0.97	0.20	0.26	0.02	0.41	0.41	0.04	0.99	0.06	0.12	0.01	0.44	0.52	0.08
PCA-MLR	0.55	0.86	1.08	0.10	0.31	1.77	0.16	0.70	0.87	1.13	0.10	0.62	1.41	0.13
PCA-ANN	0.57	0.84	1.06	0.21	1.37	1.83	0.36	0.95	0.35	0.47	0.06	1.16	1.26	0.32
PCA-SVR	0.54	0.84	1.10	0.10	1.62	1.79	0.16	0.96	0.19	0.42	0.04	1.10	1.22	0.11
PLSR-MLR	0.93	0.36	0.44	0.04	0.55	0.58	0.05	0.96	0.35	0.42	0.04	0.51	0.55	0.05
PLSR-ANN	0.94	0.33	0.38	0.08	0.25	0.31	0.06	0.94	0.41	0.49	0.06	0.71	0.78	0.20
PLSR-SVR	0.97	0.19	0.25	0.03	0.63	0.34	0.06	0.96	0.36	0.43	0.04	0.46	0.52	0.05
Ridge	0.83	0.54	0.66	0.06	1.11	1.33	0.12	0.87	0.56	0.73	0.06	0.94	1.10	0.10
LASSO	0.56	1.09	1.39	0.12	1.41	1.80	0.16	0.56	1.09	1.39	0.12	1.41	1.80	0.16
ENET	0.83	0.54	0.66	0.06	1.11	1.33	0.12	0.86	0.62	0.77	0.07	0.84	1.05	0.09
	Malda							**Uttar Dinajpur**						
SR-MLR	0.89	0.58	0.72	0.05	1.09	1.54	0.13	0.97	0.32	0.39	0.03	0.34	0.45	0.04
SR-ANN	0.93	0.29	0.39	0.08	0.62	0.65	0.06	0.98	0.21	0.30	0.03	1.66	1.66	0.28
SR-SVR	0.92	0.57	0.62	0.04	1.40	1.76	0.14	0.98	0.11	0.29	0.02	0.53	0.54	0.05
PCA-MLR	0.67	0.92	1.27	0.09	0.28	3.07	0.25	0.75	0.91	1.05	0.09	0.83	1.71	0.15
PCA-ANN	0.99	0.02	0.03	0.00	6.13	7.07	0.86	0.76	0.92	1.03	0.12	1.14	1.40	1.14
PCA-SVR	0.62	0.83	1.36	0.10	2.03	2.66	0.22	0.72	1.11	1.17	0.10	1.38	1.41	0.13
PLSR-MLR	0.90	0.57	0.70	0.05	0.61	0.81	0.07	0.98	0.24	0.29	0.02	0.39	0.47	0.04
PLSR-ANN	0.96	0.29	0.47	0.06	1.25	1.34	0.16	0.97	0.29	0.29	0.04	0.64	0.79	0.13
PLSR-SVR	0.89	0.54	0.76	0.05	0.72	0.79	0.06	0.99	0.01	0.06	0.01	0.31	0.37	0.03
Ridge	0.87	0.57	0.80	0.06	1.71	2.21	0.18	0.90	0.54	0.66	0.06	1.00	1.14	0.10
LASSO	0.56	1.09	1.39	0.12	1.41	1.80	0.16	0.56	1.09	1.39	0.12	1.41	1.80	0.16
ENET	0.80	0.77	0.97	0.07	1.71	2.43	0.20	0.87	0.59	0.76	0.06	0.87	1.16	0.10

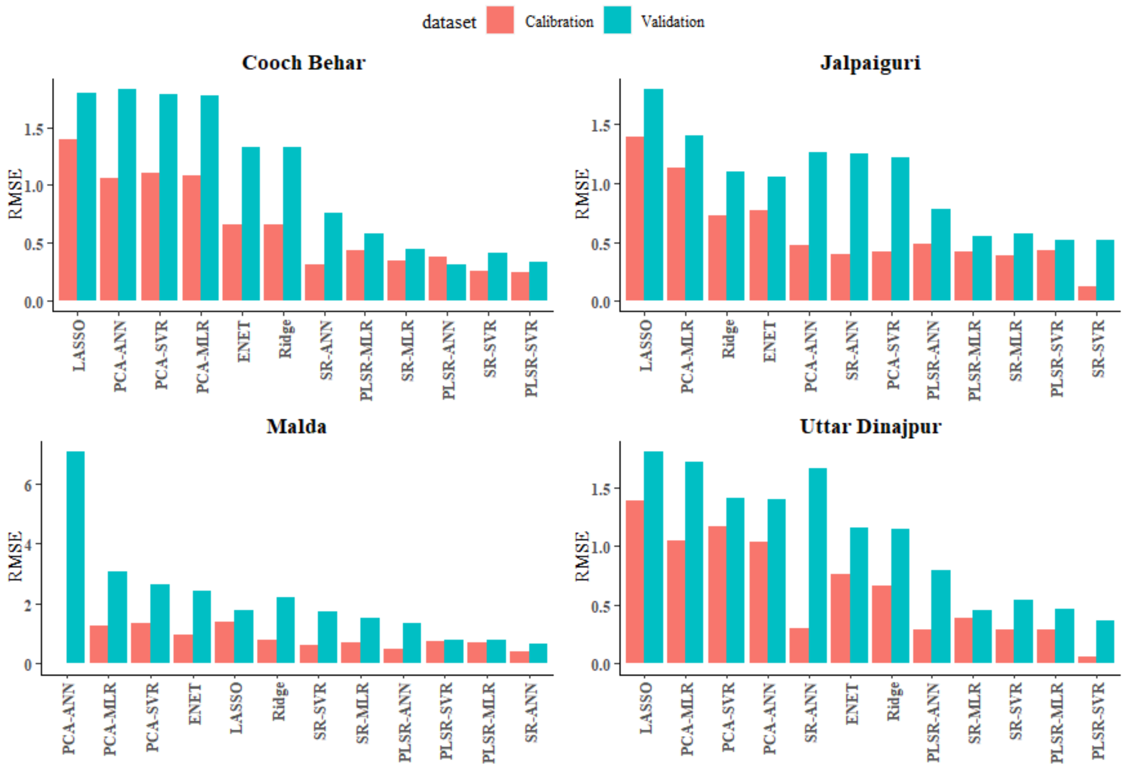

Fig. 4.20 RMSE plots of all fitted models of Jute crop of four districts

For Malda, the PCA-ANN model has the highest R^2 of 0.99. The MAE, RMSE and nRMSE are 0.02, 0.03 and zero respectively. But the error measures are very high for testing data. The SR-ANN model which has R^2 of 0.93 has MAE, RMSE and nRMSE of 0.29, 0.39 and 0.08 for training data and 0.62, 0.65 and 0.06 for testing data respectively. From Fig. 4.20, it can also be seen that RMSE of SR-ANN model during calibration as well as validation stages are low. The R^2 of PLSR-MLR and PLSR-SVR models are also high and errors are less in both training and validation stages. Hence the SR-ANN model is found to perform well for predicting the yield of jute crop in Malda followed by PLSR-MLR and PLSR-SVR models. The optimized ANN model performs better than the tradition linear models (Niazian *et al* 2018). The best-fitted SR-ANN model with Tangent hyperbolic (Tanh) activation function using weather indices selected by stepwise regression as the input is given in Fig. 4.21.

For Uttar Dinajpur, the PLSR-SVR model which has very high R^2 of 0.99 has low MAE, RMSE and nRMSE of 0.01, 0.06 and 0.01 for training data and 0.31, 0.37 and 0.03 for testing data respectively. From Fig. 4.20, it can also be seen that RMSE of PLSR-SVR model during calibration as well as validation stages are low. The PLSR-MLR, SR-SVR and SR-MLR models also have high R^2 and less error in both training and validation stages. Hence, the PLSR-SVR model is found to perform well for predicting the yield of jute crop in Uttar Dinajpur followed PLSR-MLR, SR-SVR and SR-MLR models.

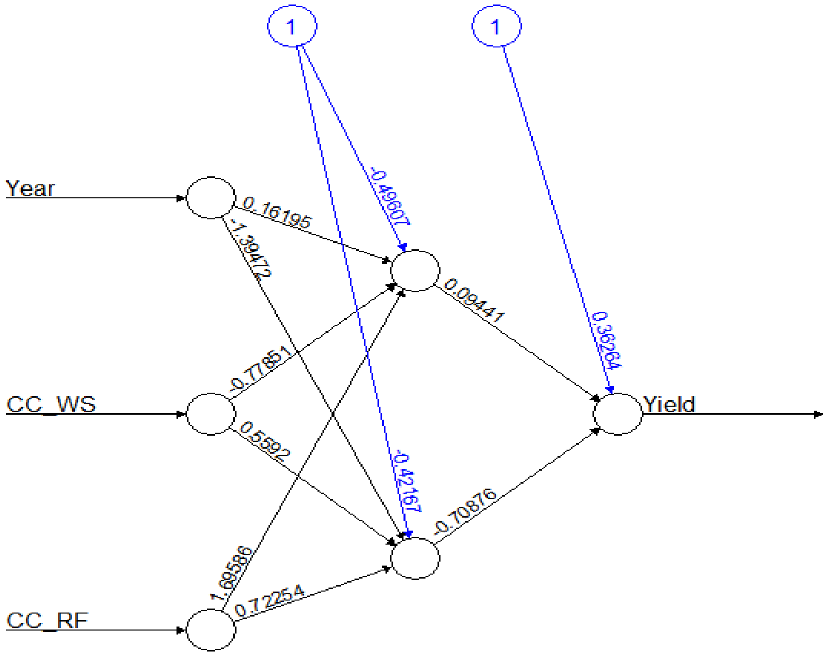

Fig. 4.21 The best fitted SR-ANN model for Jute crop of Malda district

4.5.5 Rapeseed-mustard

The model evaluation criteria of fitted models for rapeseed-mustard crop for both calibration as well as validation stages are given in the Table 4.57. For Cooch Behar, the PCA-ANN model has the highest R^2 of 0.98. The MAE, RMSE and nRMSE are 12.47, 18.47 and 0.04 respectively. But the error measures are very high for testing data which indicates the chance of overfitting. The SR-ANN model which has R^2 of 0.84 is having MAE, RMSE and nRMSE of 49.22, 58.69 and 0.12 for training data and 16.55, 18.23 and 0.09 for testing data respectively. From the Fig. 4.22, it can also be seen that RMSE of SR-ANN model during calibration as well as validation stages are low. The R^2 of PLSR-ANN model is also high level of 0.80 and errors are less in both training and validation stages. Hence, the SR-ANN model is found to perform well for predicting the yield of rapeseed-mustard crop in Cooch Behar followed by PLSR-ANN model. The ANN model using relevant input variable perform better for crop yield forecasting purposes (Emamgholizadeh *et al* 2015). The best-fitted SR-ANN model with Tangent hyperbolic (Tanh) activation function using weather indices selected by stepwise regression as the input is given in the Fig. 4.23.

Table 4.57 The model evaluation parameters in calibration and validation stages of fitted models for Rapeseed-mustard crop

Model	Calibration				Validation			Calibration				Validation		
	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE	R^2	MAE	RMSE	nRMSE	MAE	RMSE	nRMSE
	Cooch Behar							**Jalpaiguri**						
SR-MLR	0.78	57.82	69.03	0.13	78.29	91.17	0.16	0.90	31.12	42.54	0.07	38.66	41.10	0.07
SR-ANN	0.84	49.22	58.69	0.12	16.55	18.23	0.09	0.93	28.19	37.16	0.07	32.33	33.60	0.27
SR-SVR	0.83	55.97	61.85	0.12	88.73	110.59	0.20	0.94	8.61	25.13	0.04	37.99	44.23	0.07
PCA-MLR	0.71	67.33	80.29	0.15	65.23	82.57	0.14	0.82	45.03	58.63	0.09	23.46	31.24	0.05
PCA-ANN	0.98	12.47	18.47	0.04	137.08	167.22	0.79	0.94	23.87	33.05	0.06	18.56	16.75	0.16
PCA-SVR	0.69	63.25	85.12	0.16	37.02	46.79	0.08	0.78	41.17	64.12	0.10	27.98	33.47	0.06
PLSR-MLR	0.77	60.86	71.32	0.13	75.66	92.28	0.16	0.92	28.93	39.79	0.06	26.06	26.90	0.04
PLSR-ANN	0.80	56.63	65.73	0.13	23.92	24.37	0.12	0.91	32.48	41.61	0.08	29.47	37.46	0.30
PLSR-SVR	0.75	57.22	74.97	0.14	49.66	58.66	0.10	0.91	28.82	41.39	0.07	16.04	17.43	0.03
Ridge	0.68	60.42	79.37	0.15	91.87	101.91	0.19	0.91	29.16	39.31	0.06	76.38	86.51	0.13
LASSO	0.75	54.80	70.86	0.13	83.11	96.35	0.18	0.92	26.10	36.70	0.06	60.63	67.19	0.10
ENET	0.73	55.87	73.23	0.14	85.49	95.52	0.18	0.94	23.80	32.88	0.05	67.57	75.56	0.12
	Malda							**Uttar Dinajpur**						
SR-MLR	0.91	45.32	55.72	0.05	45.08	59.55	0.02	0.83	73.30	95.09	0.12	88.03	90.65	0.13
SR-ANN	0.92	43.36	53.70	0.07	89.14	97.16	0.18	0.85	69.85	88.33	0.11	96.20	116.58	0.53
SR-SVR	0.99	7.18	23.10	0.02	190.82	245.36	0.25	0.85	63.49	90.00	0.11	72.81	77.68	0.11
PCA-MLR	0.87	54.42	67.59	0.06	88.01	94.87	0.10	0.79	81.03	105.63	0.13	164.59	209.64	0.30
PCA-ANN	0.87	55.24	68.54	0.08	106.51	121.69	0.23	0.87	65.08	81.56	0.10	69.30	85.66	0.39
PCA-SVR	0.85	56.33	72.20	0.07	86.39	96.58	0.10	0.76	80.59	116.19	0.14	126.63	160.37	0.23
PLSR-MLR	0.93	41.18	59.36	0.05	38.34	44.50	0.05	0.77	81.46	109.00	0.13	128.21	151.22	0.21
PLSR-ANN	0.96	27.35	35.45	0.04	167.20	209.25	0.39	0.78	79.23	86.79	0.13	42.68	50.51	0.23
PLSR-SVR	0.93	42.08	50.24	0.05	38.30	43.14	0.04	0.76	77.28	111.75	0.14	116.98	135.37	0.19
Ridge	0.92	44.01	54.54	0.05	39.72	49.75	0.05	0.83	61.28	88.20	0.11	128.77	178.37	0.22
LASSO	0.94	39.68	49.79	0.05	45.12	50.96	0.05	0.82	67.10	91.87	0.12	116.28	134.74	0.16
ENET	0.94	39.78	49.59	0.05	46.49	51.77	0.05	0.82	65.43	90.78	0.11	109.50	136.95	0.17

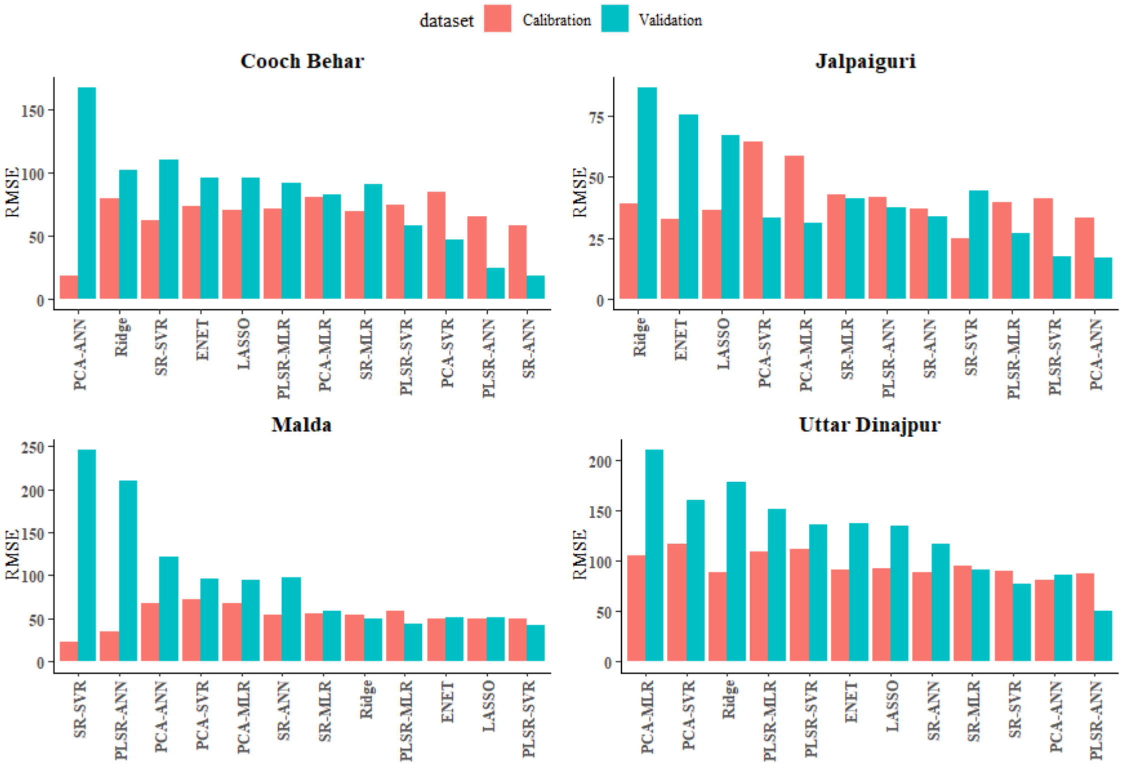

Fig. 4.22 RMSE plots of all fitted models of Rapeseed-mustard crop of four districts

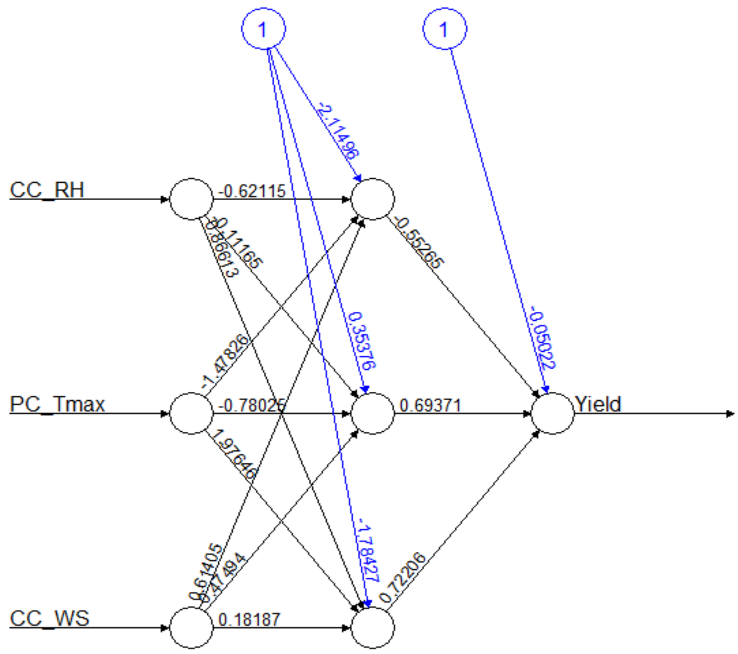

Fig. 4.23 The best fitted SR-ANN model for Rapeseed-Mustard crop of Cooch Behar

For Jalpaiguri, the PCA-ANN model which has high R^2 of 0.94 has low MAE, RMSE and nRMSE of 23.87, 33.05 and 0.06 for training data and 18.56, 16.75 and 0.16 for testing data respectively. From Fig. 4.22, it can also be seen that RMSE of PCA-ANN model during calibration as well as validation stages are low. The PLSR-SVR, PLSR-MLR and SR-SVR models also have high R^2 and less error in both the training and validation stages. Hence, the PCA-ANN model is found to perform well for predicting the yield of rapeseed-mustard crop in Jalpaiguri followed PLSR-SVR, PLSR-MLR and SR-SVR models. The best fitted PCA-ANN model with Tangent hyperbolic (Tanh) activation function using principal components as input is given in the Fig. 4.23.

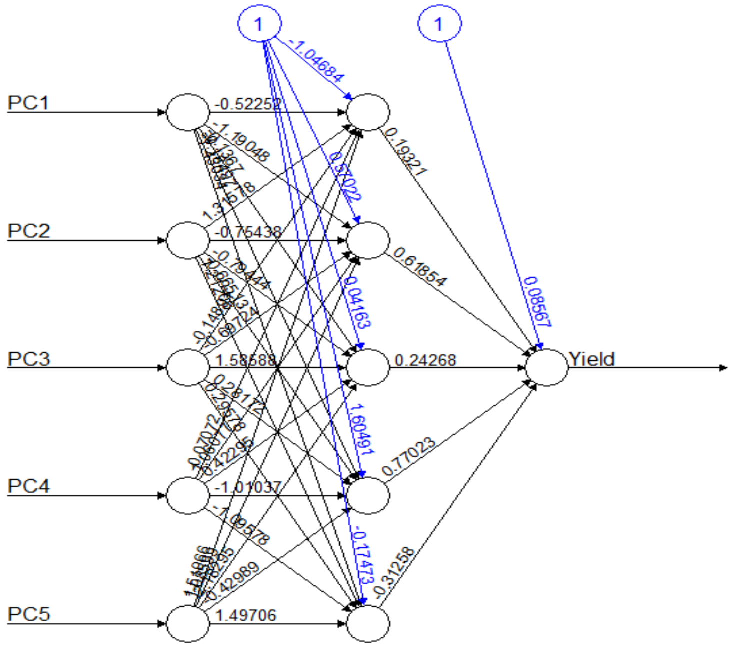

Fig. 4.23 The best fitted PCA-ANN model for Rapeseed-Mustard crop of Jalpaiguri

For Malda, the SR-SVR model has the highest R^2 of 0.99. The MAE, RMSE and nRMSE are 7.18, 23.10 and 0.02 respectively. But the error measures are very high for testing data which indicates the chance of overfitting. The PLSR-SVR model which has R^2 of 0.83 has MAE, RMSE and nRMSE of 42.08, 50.24 and 0.05 for training data and 38.30, 43.14 and 0.04 for testing data respectively. From Fig. 4.22, it can also be seen that RMSE of PLSR-SVR model during calibration as well as validation stages are low. The SR-MLR, PLSR-MLR and penalized regression models also have high R^2 and less error in both training and validation stages. Hence, the PLSR-SVR model is found to perform well for predicting the yield of rapeseed-mustard crop in Malda followed by SR-MLR, PLSR-MLR and penalized regression models.

For Uttar Dinajpur, the PLSR-ANN model which has R^2 of 0.78 has low MAE, RMSE and nRMSE of 79.23, 86.79 and 0.13 for training data and 42.68, 50.51 and 0.23 for testing data respectively. From Fig. 4.22, it can also be seen that RMSE of PLSR-ANN model during calibration as well as validation stages are low. The PCA-ANN, SR-SVR and SR-MLR models are having also high R^2 and less error in both training and validation stages. Hence, the PLSR-ANN model is found to perform well for predicting the yield of rapeseed-mustard crop in Uttar Dinajpur followed PCA-ANN, SR-SVR and SR-MLR models.

The best-fitted model as well as better better-performing model for each crop in each district are summarised in Table 4.58. While fitting regional-wise predictive models for predicting crop yield using climatic factors, the best-fitted model of a location may not perform better for other locations (Lobell and Burke, 2010).

Table 4.58 The best-fitted model for each crop in all the four districts

District	Crops	Rice	Wheat	Potato	Jute	Rapeseed-Mustard
Cooch Behar	**Best fitted model**	**Ridge model**	**PLSR-ANN**	**PLSR-ANN**	**PLSR-SVR**	**SR-ANN**
	Better performing models	ENET & LASSO	PCA-ANN	PCA-MLR & Penalized models	SR-SVR, SR-ANN, PLSR-ANN & SR-MLR	PLSR-ANN
Jalpaiguri	**Best fitted model**	**ENET model**	**SR-SVR**	**PLSR-SVR**	**SR-SVR**	**PCA-ANN**
	Better performing models	Ridge & LASSO	SR-ANN & PCA-ANN	PLSR-MLR, SR-ANN & SR-MLR	PLSR-SVR, PLSR-MLR & SR-MLR	PLSR-SVR, PLSR-MLR & SR-SVR
Malda	**Best fitted model**	**Ridge model**	**PLSR-SVR**	**PCA-SVR**	**SR-ANN**	**PLSR-SVR**
	Better performing models	ENET & LASSO	Ridge model & PLSR-MLR	SR-ANN, PCA-ANN, PLSR-ANN & SR-MLR	PLSR-MLR & PLSR-SVR	SR-MLR, PLSR-MLR & Penalized models
Uttar Dinajpur	**Best fitted model**	**ENET model**	**PLSR-SVR**	**PLSR-SVR**	**PLSR-SVR**	**PLSR-ANN**
	Better performing models	Ridge & LASSO	PLSR-MLR	PLSR-MLR, SR-SVR & SR-SVR	PLSR-MLR, SR-SVR & SR-MLR	PCA-ANN, SR-SVR & SR-MLR

The models using PLSR components as input performed better than PCA based models in most of the cases. If the dimension reduction is to fit the prediction model, the PLSR performs better than PCA (Mahesh *et al* 2018, Liu *et al* 2022). The PLSR requires fewer components than PCA (Wentzell and Montoto, 2003). Even when the number of components selected from PCA and PLSR are the same, the PLSR yields better predictive power since the PLSR components cover most of the variation in the response variable (Mevik and Wehrens, 2015). PLSR is more efficient than stepwise regression as well as PCA in identifying relevant predictors and their magnitude of influence on the response (Carrascal *et al* 2009).

4.6 Forecasting the Yield of Major Crops

The yield of all five selected crops in four districts are forecasted for the year 2020-21 using the respective best-fitted model and the same is compared with the actual yield in order to assess the forecasting performance of the best-fitted model. The actual and forecasted yield of all the five crops in four districts are given in Table 4.59.

The forecasted yield is close to the actual yield. In rice crop, the yield is overestimated by a small quantity in Jalpaiguri district and it is underestimated in the other three districts. The percent deviation of actual and forecasted yield ranges 2.52% to 3.77% which indicates that the models are performing better for forecasting the rice yield. In wheat crop, the yield is underestimated by a small quantity in Jalpaiguri district and it is overestimated in other three districts. The percent deviation of actual and forecasted yield ranges 1.42% to 4.29%.

Table 4.59 Actual and forecasted yield for the year 2020-21

District		**Rice** (kg ha^{-1})	**Wheat** (kg ha^{-1})	**Potato** (t ha^{-1})	**Jute** (bales ha^{-1})	**Rapeseed -mustard** (kg ha^{-1})
Cooch Behar	Actual	2400.00	1950.00	28.13	12.10	870.00
	Forecast	2341.00	1985.00	27.16	12.56	925.00
	Deviation	-59.00	35.00	-0.97	0.46	55.00
	Percent Deviation	**2.52**	**1.76**	**3.57**	**3.66**	**5.95**
Jalpaiguri	Actual	2640.00	2040.00	32.70	14.17	900.00
	Forecast	2735.00	1956.00	33.21	13.56	946.00
	Deviation	95.00	-84.00	0.51	-0.61	46.00
	Percent Deviation	**3.47**	**4.29**	**1.54**	**4.50**	**4.86**
Malda	Actual	2710.00	2980.00	35.68	15.50	1390.00
	Forecast	2636.00	34.23	34.23	14.20	1420.00
	Deviation	-74.00	43.00	-1.45	-0.75	30.00
	Percent Deviation	**2.81**	**1.42**	**4.24**	**5.08**	**2.11**
Uttar Dinajpur	Actual	2590.00	2640.00	29.57	13.78	1290.00
	Forecast	2496.00	2716.00	28.96	12.96	1319.00
	Deviation	-94.00	76.00	-0.61	-0.82	29.00
	Percent Deviation	**3.77**	**2.80**	**2.11**	**6.33**	**2.20**

For Potato crop, the yield is overestimated by a small quantity of 0.5 t ha^{-1} in Jalpaiguri and it is underestimated in other three districts. The percent deviation of actual and forecasted yield ranges 1.54% to 4.24%. For jute crop, the yield is overestimated by a small quantity of 0.46 bales ha^{-1} in Cooch Behar district and it is underestimated in other three districts. The percent deviation of actual and forecasted yield ranges 3.66% to

6.33%. Whereas, the yield of rapeseed-mustard is overestimated by a small quantity in all the four districts and percent deviation of actual and forecasted yield ranges 2.11% to 5.95%. The results are in agreement with results of Latwal *et al* 2019.

The crop yield can be forecasted for the future years using the observed weather data. The yield of all the five selected crops in four districts are forecasted for the years 2021-22 and 2022-23 using the respective best fitted model and the same has been given in the Table 4.60. There is a considerable increasing trend in the forecasted yield of all the five crops in four districts except wheat yield in Uttar Dinajpur and Potato yield in Malda where, slight decreasing trend is observed.

Table 4.60 The forecasted yield of five crops in four districts for 2021-22 and 2022-23

Crop	Year	Cooch Behar	Jalpaiguri	Malda	Uttar Dinajpur
Rice (kg ha^{-1})	2021-22	2640.00	2685.00	2722.00	2613.00
	2022-23	2716.00	2728.00	2765.00	2712.00
Wheat (kg ha^{-1})	2021-22	2021.00	1986.00	2942.00	2519.00
	2022-23	2062.00	2015.00	2959.00	2490.00
Potato (t ha^{-1})	2021-22	27.51	33.13	32.45	28.24
	2022-23	27.83	33.51	33.16	28.57
Jute (bales ha^{-1})	2021-22	12.25	14.10	15.61	13.21
	2022-23	12.33	14.29	15.70	13.43
Rapeseed-mustard (kg ha^{-1})	2021-22	790.00	906.00	1396.00	1312.00
	2022-23	810.00	914.00	1408.00	1323.00

The district wise yield data of major crops are obtained through the crop cutting experiments in different villages of the district which is compiled and pre-processed by Special Data Dissemination Standards (SDDS) division of Directorate of Economics and Statistics (DES). This process involves more manpower and it is cost intensive. More importantly, this complete process takes long period of time in order to get the final district wise crop yield data. The district wise yield data for the year 2020-21 is the latest available data from the authorised source. The availability of district wise yield data of the major crops for the current year is essential in order make various policy related decision for the concerned district. Crop yield forecasting is essential in order to have an idea regarding how much quantity of major crops will be produced in each district and it will be helpful for making various agriculture and policy related decisions. The yield of five major crops in four northern districts of West Bengal can be forecasted using the respective best fitted models with the use observed meteorological data.

****

Chapter – V

SUMMARY AND CONCLUSION

Chapter V

SUMMARY AND CONCLUSION

The crop yield prediction gains a growing importance for all stakeholders in agriculture starting from individual farmers to state, national and international level government as well as private organizations in order to make various agriculture-related decisions. Crop yield prediction at the regional level is essential for formulating location-specific policies. In the present study, four northern districts of West Bengal *viz.* Cooch Behar, Jalpaiguri, Malda and Uttar Dinajpur are considered. Five majorly cultivated crops *viz.* Rice, Wheat, Potato, Jute and Rapeseed-Mustard are selected for the present study.

Inter-annual crop yield variability largely depends on weather conditions that have been altered by climate change. The models based on weather variables provide a reliable forecast of crop yield. The extent of weather effect on crop yield is not only depends on the magnitude of weather factors; but also, on the distribution pattern of each weather factor over the crop season. Hence, it is required to give appropriate weightage to weekly weather conditions according to its effect on ultimate crop yield. In order to give weightage to the respective week's weather conditions, correlation-based weighted indices have been utilized by many researchers in the past four decades. The path coefficient calculated from path analysis measures only the direct effect of an independent variable on the ultimate response. Hence, path coefficient-based weighted index is proposed in this study in a view of getting more precise and accurate weighted weather indices. The weather parameters considered for the study are maximum temperature, minimum temperature, relative humidity, rainfall and windspeed.

The study area is examined for climate suitability for the cultivation of crops using agrometeorological indices. The agrometeorological indices such as Growing Degree Days (GDD), Diurnal Temperature Range (DTR), Relative Temperature Disparity (RTD) and Number of Rainy Days have been calculated for each crop in each district using the weather data of the period in which the crop is grown. The GDD of winter crops are high followed by rice and jute season in all four districts. The higher GDD in the winter season than the other two seasons can be attributed to the low threshold temperature requirement of winter crop is low. The GDD of Malda is high for winter crops as well as rice and jute crops followed by Uttar Dinajpur and Jalpaiguri. The GDD of Cooch Behar is comparatively low. There is a significant increasing trend in the GDD of Uttar Dinajpur, Cooch Behar and Malda districts during aman rice crop season. The average DTR and RTD is high in the winter season followed by jute and aman rice season. There is an increasing trend in the DTR of Cooch Behar and Jalpaiguri districts and decreasing trend in Malda and Uttar Dinajpur in all three seasons. The number of rainy days is higher during rice season due to the coincidence of monsoon followed by jute season due to summer showers. The number of rainy days during winter is low. There is a significant decreasing trend in the number of rainy days in Malda during the winter season and there is a significant decreasing trend in Cooch Behar during aman rice season.

The weekly weather data of the period in which a particular crop is grown are used to calculate the unweighted and weighted weather indices. The unweighted indices are calculated as the average of weekly weather data of all the weather parameters except rainfall for which the cumulative value of the crop growing period is calculated instead of average.

Before calculating the correlation coefficient-based as well as path coefficient-based weighted indices, the yield of each crop in each district is tested for the presence of trend. Modified Mann Kendall is employed to detect the presence of a significant trend in the yield. There is a significant trend in the yield of all the five crops in all four districts except rapeseed-mustard yield in Cooch Behar. The significant increasing trend in crop yield can be attributed to the introduction high high-yielding varieties, improved cultivational practices *etc.* over the years. The respective yield is detrended before the calculation of weighted weather indices. Detrending of yield before calculating correlation as well as path coefficient is done with an aim to study the actual effect of weather parameters on yield by removing trend-causing factors. The detrending of yield has been done by subtracting the predicted yield from actual yield where the predicted yield is obtained by regressing the yield using year number as the explanatory variable.

Five unweighted, five correlation coefficients and five path coefficient-based weighted indices have been calculated for each crop in each district as there are five weather parameters. Both the weighted indices of the majority of the weather parameters are lesser than weighted indices in most of the cases and some of the weighted indices are negative due to the negative association of the weather parameter on yield. Hence, the weighted indices differ from unweighted indices in terms of weighting a particular weather parameter based on its relationship to yield. The correlation-based weighted indices are weighting a particular weather parameter according to its correlation with yield, while path coefficient-based weighted indices weighting a particular weather parameter according to its direct effect on yield. Whereas, the unweighted indices are just an average of the weather parameter during the weeks in which the particular crop is grown.

Since the trend in yield is significant in almost all crops in every district, the year number is also used as a trend variable in the model to account for the trend-causing factor along with weather indices. Therefore, there are sixteen input or explanatory variables for each crop in each district. The inclusion of all the explanatory variables in the model leads to a complex model which necessitates the estimation of many parameter values that leads to the slow learning of the model. In some cases, the inclusion of some of the redundant or nonsignificant variables into the model leads to the overfitting issue. With this aim, three variable selection or dimension reduction methodologies *viz.* Stepwise Regression (SR), Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR) are employed.

From the results of stepwise regression, it can be observed that only two to seven explanatory variables are selected. However, the information present in the other variables is completely ignored. In order to achieve variable selection as well as to store maximum information present in the entire data in a few of the selected variables simultaneously, two dimension reduction techniques *viz.* Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR) are employed.

Principal Component Analysis (PCA) is conducted for each crop in all four districts by using their respective weather indices as well as time trend variables corresponding to the particular crop of a district. Sixteen Principal Components (PCs) are derived for each crop in each district as there are sixteen variables. The decision on the number of components to be selected is taken based on three criteria *viz.* cumulative percent variability explained, eigenvalue criteria and scree plot. The number of principal components selected ranges between four to six.

Similarly, Partial Least Square Regression (PLSR) has been carried out for each crop of all four districts by taking crop yield as the dependent variable and their respective weather indices as well as time trend variable as independent variables. Sixteen partial least square components are obtained for each crop in each district as there are sixteen independent variables. The optimum number of components is selected using a cross-validation technique. The optimum number of PLSR components ranges between one to seven. Since the PLSR components are derived by maximizing the covariance between the yield and explanatory variables set, the derived PLSR components are expected to have maximum variability in the yield.

Since crop yield is complexly associated with many environmental factors, different models with different functional forms have to be explored in order to obtain location-specific best-performing models. In this present study, the performance of Multiple Linear Regression (MLR), three penalized regression models *viz.* Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) and two machine learning models *viz.* Artificial Neural Network (ANN) and Support Vector Regression (SVR) have been evaluated.

The MLR, ANN and SVR models have been fitted under three variable selection conditions. Since the penalized regression models are doing inner variable selection by shrinking the coefficient values, there is no necessity for variable selection or dimension reduction. Hence, twelve models have been fitted for each crop in each district. 80% of the data are randomly selected for model training and the remaining 20% of data are utilized for validation of the fitted models. The random selection is employed in order to ensure the presence of recent as well as past year's data in both training and testing data set. The performance of the fitted models has been evaluated using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and normalized Root Mean Square Error (nRMSE). The model that performs better in both the training and validation stages is considered the best-fitted model. The best-fitted model for each crop in each district is selected.

From the results of MLR model, it can be observed that most of PLSR components are significantly influencing the yield of the respective crop in all four districts in PLSR-MLR models. The earlier PLSR components are significantly influencing the response variable, since the PLSR components are derived by considering the response variable also. Only a few last components are non-significant. But in PCA based model there is no such order.

Three ANN models have been fitted for each crop in each district by taking indices selected from stepwise regression as input (SR-ANN), principal components as input (PCA-ANN) and PLSR components as input (PLSR-ANN). The Multilayer Perceptron (MLP) architecture of the ANN model is trained using the Backpropagation (BP) algorithm. The performance of three activation functions *viz.* Logistic, Tangent Hyperbolic (Tanh) and Restricted Linear Unit (ReLU) have been evaluated. The number of input layer neurons of SR-ANN, PCA-ANN and PLSR-ANN models are the number of indices selected using SR, the number of principal components selected from PCA and the number of components selected from PLSR respectively. The number of output layer neurons is one as there is only one output variable which is crop yield. The number of neurons in the hidden layer is determined using 10-fold cross-validation technique.

The optimum number of hidden layer neurons is less than or equal to the number of input layer neurons. More number of hidden layer neurons cause slow learning as well as overfitting issues and a smaller number of neurons may be insufficient. The ANN model with the optimum number of hidden layer neurons performs better. The optimum learning rate ranges between 0.02 and 0.10. The ANN models fitted by optimizing critical aspects such as learning rate and number of hidden layer neurons increase the predictive power of the crop yield prediction model. The ANN model using Tanh activation function can converge to the lowest possible error with a reasonable learning rate. Hence, the Tanh function is found to perform better as a hidden layer activation function for predicting crop yield.

Three Support Vector Regression (SVR) models have been fitted to each crop in each district by taking indices selected from stepwise regression as input (SR-SVR), principal components as input (PCA-SVR) and PLSR components as input (PLSR-SVR). The performance of SVR model under a Linear kernel as well as two nonlinear kernels *viz.* Polynomial and Radial Basis Function (RBF) have been evaluated. The linear kernel has only two hyperparameters, cost and epsilon. RBF kernel has an additional parameter gamma. The range of cost (C) parameter is set as 1 to 32. The range of epsilon (ε) and gamma (γ) are set from 0 to 1 with 0.01 interval. The degree of polynomial (d) is the fourth parameter in the polynomial kernel which is tuned from 1 to 10. The best combination of these hyperparameters is selected for each kernels using a grid search algorithm based on a 10-fold cross-validation technique. The SVR model using RBF kernel function achieved the lowest RMSE in most of all the cases than the other two kernels. The hyperparameter-tuned SVR using RBF kernel was found to perform better for predicting the crop yield. The grid search algorithm is effective for optimizing the SVR hyperparameters.

The performance of three penalized regression models viz. Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) have been evaluated for each crop in each district. For fitting the penalized regression modes, the penalty parameters Alpha (α) and Lambda (λ) have to be optimized. The alpha value for Ridge and LASSO regression is fixed; that are 0 and 1 respectively. Hence only lambda (λ) value has to be optimized for Ridge and LASSO models. For ENET model, both alpha (α) and lambda (λ) have to be optimized. The penalty parameters have been optimized using 10-fold cross-validation. In comparison with OLS estimates, the regression coefficients of each index are shrunken in all three penalized regression models. In Ridge regression, the coefficient values are never shrunken to zero. But in LASSO and ENET models, some of the coefficients are shrunken to zero, thereby variable selection is accomplished. Only important variables are retained in the final LASSO and ENET models. In most of the cases, unweighted indices are removed. The correlation and path coefficient-based indices are retained in the final model. Due to the coefficient shrinkage as well as variable selection in penalized regression, it is expected that penalized regression models perform better.

The penalized regression models provide stable performance in both training as well as validation stages for rice crop. The Ridge model is found to perform better for Cooch Behar and Malda districts and ENET model is found to perform well for Jalpaiguri and Uttar Dinajpur districts.

For other crops, the machine learning models accompanied with PLSR components as input is found to perform well in both training and validation stages. The PLSR-SVR is found to perform better in seven out of sixteen cases of wheat, potato, jute and rapeseed-mustard crops of four districts. The SR-SVR and PCA-SVR performed well in two and one cases respectively. The SR-ANN, PCA-ANN and PLSR-ANN models are performed better in two cases each.

The path coefficient-based weighted indices are found to have a significant effect on yield in stepwise regression. The path coefficient-based weighted indices are retained in the final penalized regression models as well. Hence, the proposed path coefficient-based weighted indices are having a predictable effect on crop yield along with the existing unweighted and correlation coefficient-based weighted indices.

The models using PLSR components as input performed better than PCA-based models in most of the cases. If the dimension reduction is to fit the prediction model, the PLSR performs better than PCA. Even when the number of components selected from PCA and PLSR are the same, the PLSR yields better predictive power since the PLSR components cover most of the variation in the response variable. PLSR is more efficient than stepwise regression as well as PCA in identifying relevant predictors and their magnitude of influence on the response. The machine learning models that are able to capture complex nonlinear associateship between weather variables and crop yield outperformed the traditional linear regression model.

The yield of all five selected crops in four districts is forecasted for the year 2020-21 using the respective best-fitted model and the same is compared with the actual yield in order to assess the forecasting performance of the best-fitted model. It is found that the forecasted yield is close to the actual yield. The percent deviation ranges between 1.54% to 6.33%. Hence, the best-fitted models perform better for forecasting the yield. The yield of all the five selected crops in four districts is forecasted for the years 2021-22 and 2022-23 using the respective best-fitted model. The yield of five major crops in four northern districts of West Bengal can be forecasted using the respective best-fitted models with the use of observed weather data.

****

Chapter – VI

FUTURE SCOPE OF RESEARCH

Chapter VI

FUTURE SCOPE OF RESEARCH

- ❖ The methodologies used in this research can also be applied to other crops in this region as well as other areas.
- ❖ The weather factor which has close relation with crop yield such as Photosynthetically Active Radiation (PAR), Bright Sunshine Hours (BSH) can also be used. But the non-availability of the long-term data on these parameters is a constraint.
- ❖ The data on plant characters, pest and disease incidence and soil nutrient availability can also be incorporated along with weather factors while fitting models to predict crop yield.
- ❖ The statistical and machine learning models can also be fitted using agrotechnological indices as input variables for predicting crop yield and the performance of the same can be compared with the models based on weather indices.
- ❖ Since the LASSO and ENET models are also doing variable selection, the variables selected from the above can be used to fit machine learning models.
- ❖ Machine learning based feature selection techniques can be employed and the same can be compared with Stepwise Regression (SR), Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR).
- ❖ The performance of ANN model using other activation functions and SVR model with other kernel functions can also be evaluated.
- ❖ The advanced machine learning models such as Decision Tree, Random Forest, Extreme Gradient Boost algorithm and deep learning models such as Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) can be explored to predict the crop yield.

****

REFERENCES

REFERENCES

Abbas F, Afzaal H, Farooque A A and Tang S (2020) Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy* **10(7)**: 1046-1062.

Abbas S, Kousar S and Khan M S (2022) The role of climate change in food security; empirical evidence over Punjab regions, Pakistan. *Environmental Science and Pollution Research* **29(35)**: 53718-53736.

Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley interdisciplinary reviews: computational statistics* **2(1)**: 97-106.

Abdi H (2003) Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences* **6(4)**: 792-795.

Abhinaya D, Patil S G, Ga D, Djanaguiraman M and Raj A S (2021) Use of Statistical Models in Predicting Groundnut Yield in Relation to Weather Parameters. *Madras Agricultural Journal* **108(special)**: 107-104.

Agrawal R, Jain R C and Jha M P (1983) Joint effects of weather variables on rice yield. *Mausam* **34(2)**: 189-194.

Aggarwal C C (2018) Neural networks and deep learning. *Springer* **10(978)**: 497.

Aggarwal P K (2008) Global climate change and Indian agriculture: impacts, adaptation and mitigation. *Indian Journal of Agricultural Sciences* **78(11)**: 911-919.

Agrawal R, Jain R C and Jha M P (1986) Models for studying rice crop weather relationship. *Mausam* **37(1)**: 67-70.

Agrawal R and Mehta S (2007) Weather based forecasting of crop yield, pests and diseases- IASRI models. *Journal of Indian Society of Agricultural Statistics* **61(2)**: 255-263.

Agustina S D, Bella C and Ramadhan M A (2018) Support vector regression algorithm modeling to predict the availability of foodstuff in Indonesia to face the demographic bonus. *In. Journal of Physics: Conference Series* **1028(1)**: 012240.

Ajith S, Debnath M K, Gupta D S, Basak P, Bandyopadyay S, Kheroar, S and Ragini H R (2023) Comparative evaluation of penalized regression models with multiple linear regression for predicting rapeseed-mustard yield: Weather-indices based approach. *Journal of Agrometeorology* **25(3):** 432-439.

Akhter J, Mandal R, Chattopadhyay R, Joseph S, Dey A, Nageswararao M M, Pattanaik D R and Sahai A K (2021) Kharif rice yield prediction over Gangetic West Bengal using IITM-IMD extended range forecast products. *Theoretical and Applied Climatology* **145(3-4)**: 1089-1100.

Aktar N (2015) Agricultural productivity and productivity regions in West Bengal. *The NEHU Journal* **8(2)**: 49-61.

Al-Azies H, Trishnanti D and PH E M (2019) Comparison of kernel support vector machine (SVM) in classification of human development index (HDI). *IPTEK Journal of Proceedings Series* **30(6)**: 53-57.

Alkaff M, Khatimi H, Puspita W and Sari Y (2019) Modelling and predicting wetland rice production using support vector regression. *Telecommunication Computing Electronics and Control* **17(2)**: 819-825.

Alrefaee S D, Al-Bakal S M and Algamal Z Y (2022) Hyperparameters optimization of support vector regression using black hole algorithm. *International Journal of Nonlinear Analysis and Applications* **13(1)**:3441-3450.

Alwin D F and Hauser R M (1975) The decomposition of effects in path analysis. *American sociological review* **1**: 37-47.

Amaratunga V, Wickramasinghe L, Perera A, Jayasinghe J and Rathnayake U (2020) Artificial neural network to estimate the paddy yield prediction using climatic data. *Mathematical Problems in Engineering* **18** :1-11.

Ansarifar J, Wang L and Archontoulis S V (2021) An interaction regression model for crop yield prediction. *Scientific reports* **11(1)**: 17754.

Aravind K S, Vashisth A, Krishanan P and Das B (2022) Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *Journal of Agrometeorology* **24(1)**: 18-25.

Artigue H and Smith G (2019) The principal problem with principal components regression. *Cogent Mathematics & Statistics* **6(1)**: 1622190.

Arun D P, Malik J S and Kumari N (2022) *Impact of Climate Change on Agriculture: An Overview*. pp 1-10. New Delhi Publishers, New Delhi.

Asseng S, Zhu Y, Wang E and Zhang W (2015) Crop modeling for climate change impact and adaptation. *Crop physiology* **2015**: 505-546.

Awad M and Khanna R (2015) *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. pp 67-80. Apress Berkeley, CA.

Azfar M, Sisodia B V, Rai V N and Devi M (2015) Pre-harvest forecast models for rapeseed & mustard yield using principal component analysis of weather variables. *Mausam* **66(4)**: 761-766.

Bahrami M, Shabani A, Mahmoudi M R and Didari S (2020) Determination of effective weather parameters on rainfed wheat yield using backward multiple linear regressions based on relative importance metrics. *Complexity* **2020**: 1-10.

Bal S K, Mukherjee J, Mallick K and Hundal S S (2004) Wheat yield forecasting models for Ludhiana district of Punjab state. *Journal of Agrometeorology* **6(special)**: 161-165.

Banakara K B, Pandya H R and Garde Y A (2019) Pre-harvest forecast of kharif rice yield using PCA and MLR technique in Navsari district of Gujarat. *Journal of Agrometeorology* **21(3)**: 336-343.

Banakara K B, Pandya H, Garde Y A and Ojha S (2017) Pre-harvest crop modelling of kharif rice using weather parameters in Valsad district of south Gujarat. *Editorial Board* **2017**: 196-200.

Barma P, Tarafder H K and Moktan R S (2021) Performance of different mustard varieties in terai zone of West Bengal. *International Journal of Plant Sciences* **16(1)**: 31-33.

Barman D and Anoop M (2023) Assessment of Efficiency of Marketing Channels and Constraints in Marketing of Jute in Cooch Behar District, West Bengal. *Economic Affairs* **68(1)**: 417-421.

Barman D, Kundu D, Ghorai A and Mitra S (2014) Determination of evapotranspiration and crop coefficient of tossa jute (Corchorus olitorius). *Journal of Agricultural Physics* **14(1)**: 67-72

Basha S M, Rajput D S, Janet J, Somula R S and Ram S (2020) Principles and practices of making agriculture sustainable: crop yield prediction using Random Forest. *Scalable Computing: Practice and Experience* **21(4)**: 591-599.

Batool D, Shahbaz M, Shahzad Asif H, Shaukat K, Alam T M, Hameed I A, Ramzan Z, Waheed A, Aljuaid H and Luo S (2022) A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning. *Plants* **11(15)**: 1925.

Bauer J O and Drabant B (2021) Principal loading analysis. *Journal of Multivariate Analysis* **184**: 104754.

Bazgir S, Kamali G and Mortazavi A 2007 Wheat yield prediction through agrometeorological indices for Hamedan, Iran. *BIABAN (Desert Journal)* **12**: 33-38

Bhardwaj M, Kumar P, Kumar S, Dagar V and 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. *Environmental Science and Pollution Research* **29(21)**: 31861-31885.

Bhat S and Vidya R (2018) Partial Least Squares and Other Biased Regression Methods. *Thailand Statistician* **16(1)**: 38-55.

Birthal P S, Khan T, Negi D S and Agarwal S (2014) Impact of climate change on yields of major food crops in India: Implications for food security. *Agricultural Economics Research Review* **27(2)**:145-155.

Boneh S and Mendieta G R (1994) Variable selection in regression models using principal components. *Communications in Statistics-Theory and Methods* **23(1)**: 197-213.

Boomiraj K, Chakrabarti B, Aggarwal P K, Choudhary R and Chander S (2010) Assessing the vulnerability of Indian mustard to climate change. *Agriculture, ecosystems & environment* **138(3-4)**: 265-273.

Brereton R G and Lloyd G R (2010) Support vector machines for classification and regression. *Analyst* **135(2)**: 230-267.

Cangelosi R and Goriely A (2007) Component retention in principal component analysis with application to cDNA microarray data. *Biology direct* **2(1)**: 1-21.

Cardoso A O, Pinto H S, de Ávila A M, da Silva Dias P L, Marin F R and Pilau F (2010) Extended time weather forecasts contributes to agricultural productivity estimates. *Theoretical and applied climatology* **102**: 343-350.

Carrascal L M, Galván I and Gordo O (2009) Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* **118(5)**: 681-690.

Carter H R, Fomby T B and Johnson S R (1997) Component selection norms for principal components regression. *Communications in Statistics-Theory and Methods* **6(4)**: 309-334.

Cattell R B (1966) The scree test for the number of factors. *Multivariate behavioral research* **1(2)**: 245-276.

Chandio A A, Nasereldin Y A, Anh D L, Tang Y, Sargani G R and Zhang H (2022) The impact of technological progress and climate change on food crop production: Evidence from Sichuan—China. *International Journal of Environmental Research and Public Health* **19(16)**: 9863.

Chandrasiri C K, Tsusaka T W, Ho T D, Zulfiqar F and Datta A (2023) Impacts of climate change on paddy yields in different climatic zones of Sri Lanka: a panel data approach. *Asia-Pacific Journal of Regional Science* **7(2)**: 455-489.

Charalampopoulos I (2021) Agrometeorological Conditions and Agroclimatic Trends for the Maize and Wheat Crops in the Balkan Region. *Atmosphere* **12(6)**: 671.

Chatterjee S, Khan A, Akbari H and Wang Y (2016) Monotonic trends in spatio-temporal distribution and concentration of monsoon precipitation (1901–2002), West Bengal, India. *Atmospheric Research* **182**: 54-75.

Chatterjee S, Ray M, Halder P and Goswami R (2013) Economic characterization of predominant farming systems in West Bengal, India. *American Journal of Agriculture and Forestry* **1(3)**: 40-47.

Chaturvedi D K (2008) Factors affecting the performance of artificial neural network models. *Soft Computing: Techniques and its Applications in Electrical Engineering* **103**: 51-85.

Chaudhari N V, Kumar N, Parmar P K, Dakhore K K, Chaudhari S N and Chandrawanshi S K (2019) Thermal indices in relation to crop phenology and yield of rice (Oryza sativa L.) grown in the south Gujarat region. *Journal of Pharmacognosy and Phytochemistry* **8(2)**: 146-149.

Chen H, Wu W and Liu H B (2016) Assessing the relative importance of climate variables to rice yield variation using support vector machines. *Theoretical and Applied Climatology* **126**: 105-111.

Cheng B and Wu X (2006) A modified PLSR method in prediction. *Journal of Data Science* **4(3)**: 257-274.

Chutia S, Deka R L, Goswami J and Phukon M H (2021) Forecasting rice yield through modified Hendrick and Scholl technique in the Brahmaputra valley of Assam. *Journal of Agrometeorology* **23(1)**: 106-112.

Cook R D and Forzani L (2020) Envelopes: A new chapter in partial least squares regression. *Journal of Chemometrics* **34(10)**: e3287.

Cortes C and Vapnik V (1995) Support vector networks. *Machine Learning* **20**: 273–297.

Dahikar S S and Rode S V (2014) Agricultural crop yield prediction using artificial neural network approach. *International journal of innovative research in electrical, electronics, instrumentation and control engineering* **2(1):** 683-286.

Dar A N (2021) Principal Component Analysis (PCA). *Global Scientific Journal* **9(7)**: 242-250.

Das A and Kumar S (2018) Growth performance of rice in west Bengal agriculture: a spatio temporal analysis. *Economic affairs* **63(4)**: 897-903.

Das B, Nair B, Reddy V K and 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. *International Journal of Biometeorology* **62(10)**: 1809-1822.

Das B, Nair B, Arunachalam V, Reddy K V, Venkatesh P, Chakraborty D and Desai S (2020) Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the west coast of India. *International Journal of Biometeorology* **64**: 1111-1123.

Das J, Mandal T and Saha P (2019) Spatio-temporal trend and change point detection of winter temperature of North Bengal, India. *Spatial Information Research* **27**: 411-424.

Das S and Kumar M (2019) Development of Yield Forecasting Model for Wheat by using Weather Variables. *Research Journal of Agricultural Science* **10(3)**: 534-536.

Datta P and Das S (2019) Analysis of long-term seasonal and annual temperature trends in North Bengal, India. *Spatial Information Research* **27(4)**: 475-496.

Daoud J I (2017) Multicollinearity and regression analysis. *In Journal of Physics: Conference Series* **949(1)**: 1-6.

Dhamija V, Shukla R, Gornott C and Joshi P K (2020) Consistency in vulnerability assessments of wheat to climate change—A district-level analysis in India. *Sustainability* **12(19)**: 8256.

Delerce S, Dorado H, Grillon A, Rebolledo M C, Prager S D, Patiño V H, Garcés Varón G and Jiménez D (2016) Assessing weather-yield relationships in rice at local scale using data mining approaches. *PloS one* **11(8)**: 1-25.

Dewey D R and Lu K (1959) A correlation and path-coefficient analysis of components of crested wheatgrass seed production. *Agronomy journal* **51(9)**: 515-518.

Dhekale B S, Sawant P K and Upadhye T (2014) Weather based pre-harvest forecasting of rice at Kolhapur (Maharashtra). *Trends Bioscience* **7**: 39-41.

Dorugade A V and Kashid D N (2010) Alternative method for choosing ridge parameter for regression. *Applied Mathematical Sciences* **4(9)**: 447-456.

Duan L, Xie H, Li Z, Yuan H, Guo Y, Xiao X and Zhou Q (2020) Use of partial least squares regression to identify factors controlling rice yield in Southern China. *Agronomy Journal* **112(3)**: 1502-1516.

Dubey R, Pathak H, Chakrabarti B, Singh S, Gupta D K and Harit R C (2020) Impact of terminal heat stress on wheat yield in India and options for adaptation. *Agricultural Systems* **181**: 102826.

Duffy J R, Watt J and Duffy R J (1981) Path analysis: A strategy for investigating multivariate causal relationships in communication disorders. *Journal of Speech, Language, and Hearing Research* **24(4)**: 474-490.

Emamgholizadeh S, Parsaeian Malda and Baradaran M (2015) Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy* **68**: 89-96.

Eslamian S, Ghasemizadeh M, Biabanaki M and Talebizadeh M (2010) A principal component regression method for estimating low flow index. *Water Resources Management* **24**: 2553-2566.

Fahrmeir L, Kneib T, Lang S and Marx B (2013) *Regression: Models, Methods and Applications*. pp 21-72. Springer Science & Business Media.

Farrar D E and Glauber R R (1967) Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics* **1**: 92-107.

Feng J and Lu S (2019) Performance analysis of various activation functions in artificial neural networks. *In Journal of physics: conference series* **1237(2)**: 022030.

Fisher R A (1924) The influence of rainfall on the yield of wheat at Rothamsted. *Royal Society (London)* **213**: 89-142.

Firinguetti L, Kibria G and Araya R (2017) Study of partial least squares and ridge regression methods. *Communications in Statistics-Simulation and Computation* **46(8)**: 6631-6644.

Freedman D A (1987) As others see us: A case study in path analysis. *Journal of educational statistics* **12(2)**: 101-128.

Gadedjisso-Tossou A, Adjegan K I and Kablan A K (2021) Rainfall and temperature trend analysis by Mann–Kendall test and significance for Rainfed Cereal Yields in Northern Togo. *Science* **3(1)**: 17.

Garde Y A, Dhekale B S and Singh S (2015) Different approaches on pre harvest forecasting of wheat yield. *Journal of Applied and Natural Science* **7(2)**: 839-843.

Gardner M W and Dorling S R (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment* **32(14-15)**: 2627-2636.

Gareth J, Daniela W, Trevor H and Robert T (2013) *An introduction to statistical learning: with applications in R*. Spinger.

Ghani I M and Ahmad S (2010) Stepwise multiple regression method to forecast fish landing. *Procedia-Social and Behavioral Sciences* **8**: 549-554.

Ghosh S, Mahato K, Gorain S, Das U and Mondal B (2017) Resilience of agriculture reducing vulnerability to climate change in West Bengal. *Current Advances in Agricultural Sciences* **9(2)**: 170-177

Goktas A and Akkus O (2020) Comparison of partial least squares with other prediction methods via generated data. *Journal of Statistical Computation and Simulation* **90(16)**: 3009-3024.

Gornott C and Wechsung F (2016) Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany. *Agricultural and Forest Meteorology* **217**: 89-100.

Goyal A and Vashisth A (2021) Mustard yield prediction using machine learning approach. *Journal of Agricultural Physics* **21(2)**: 445-456.

Goyal M (2018) Pre-harvest wheat yield forecast through agro-meteorological indices for northern region of Haryana. *International Journal of Applied Mathematics Statistical Sciences* **7(3)**: 29-40

Guenther N and Schonlau M (2016) Support vector machines. *The Stata Journal* **16(4)**: 917-937.

Guo N, Gui W, Chen W, Tian X, Qiu W, Tian Z and Zhang X (2020) Using improved support vector regression to predict the transmitted energy consumption data by distributed wireless sensor network. *EURASIP Journal on Wireless Communications and Networking* **(1)**:1-6.

Guo Y, Fu Y, Hao F, Zhang X, Wu W, Jin X, Bryant C R and Senthilnath J (2021) Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecological Indicators* **120**: 106935.

Guo Y, Wu W, Du M, Bryant C R, Li Y, Wang Y and Huang H (2019) Assessing potential climate change impacts and adaptive measures on rice yields: The case of zhejiang province in China. *Sustainability* **11(8)**: 2372-2347.

Gupta S, Singh A, Kumar A, Shahi U P, Sinha N K and Roy S (2018) Yield forecasting of wheat and mustard for western Uttar Pradesh using statistical model. *Journal of Agrometeorology* **20(1)**: 66-68.

Habib-ur-Rahman M, Ahmad A, Raza A, Hasnain M U, Alharby H F, Alzahrani Y M, Bamagoos A A, Hakeem K R, Ahmad S, Nasim W and Ali S (2022) Impact of climate change on agricultural production; Issues, challenges, and opportunities in Asia. *Frontiers in Plant Science* **13**: 925548.

Hadi A S and Ling RF (1998) Some cautionary notes on the use of principal components regression. *The American Statistician* **52(1)**: 15-19

Halder P, Dey R K and Mandal S (2022) Long period trend analysis of annual and seasonal rainfall in West Bengal, India (1901-2020). *Research square* **2022**: 1-25.

Hamed K H and Rao A R (1998) A modified Mann-Kendall trend test for autocorrelated data. *Journal of hydrology* **204(1-4)**: 182-196.

Han I S, Kim M, Lee C H, Cha W, Ham B K, Jeong J H, Lee H, Chung C B and Han C (2003) Application of partial least squares methods to a terephthalic acid manufacturing process for product quality control. *Korean Journal of Chemical Engineering* **20**: 977-984.

Han X, Liu F, He X and Ling F (2022) Research on rice yield prediction model based on deep learning. *Computational Intelligence and Neuroscience*: 1-9.

Hao Z, Geng X, Wang F and Zheng J (2018) Impacts of climate change on agrometeorological indices at winter wheat overwintering stage in Northern China during 2021–2050. *International Journal of Climatology* **38(15)**: 5576-5588.

Hasan M A, Xu S, Kabir M M and Ahmad S (2016) Performance evaluation of different kernels for support vector machine used in intrusion detection system. *International Journal of Computer Networks & Communications* **8(6)**: 39-53.

Henderson D A and Denison D R (1989) Stepwise regression in social and psychological research. *Psychological Reports* **64(1)**: 251-257.

Hendricks W A and Scholl J C (1941) Techniques in measuring joint relationships: The joint effects of temperature and precipitation on corn yield. *North Carolina Agricultural Experiment Station Research Bulletin* **26(74)**: 35-46.

Holland S M (2019) *Principal components analysis (PCA)*. pp 1-11. Department of Geology, University of Georgia, Athens, GA.

Hope K (1971) Path analysis: supplementary procedures. *Sociology* **5(2)**: 225-241.

Hoskuldsson A (1988) PLS regression methods. *Journal of chemometrics* **2(3)**: 211-128.

Hotelling H (1957) The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology* **10(2)**: 69-79.

Hoogenboom G (2000) Contribution of agrometeorology to the simulation of crop production and its applications. *Agricultural and forest meteorology* **103(1-2)**: 137-157.

Hong H, Li F W and Xu J (2019) Climate risks and market efficiency. *Journal of econometrics* **208(1)**: 265-281.

Hu Y, Wei X, Hao M, Fu W, Zhao J and Wang Z (2018) Partial least squares regression for determining factors controlling winter wheat yield. *Agronomy Journal* **110(1)**: 281-292.

Huang N, Song Y, Wang J, Zhang Z, Ma S, Jiang K and Pan Z (2022) Climatic threshold of crop production and climate change adaptation: A case of winter wheat production in China. *Frontiers in Ecology and Evolution* **10**: 1019436.

Huberty C J (1989) Problems with stepwise methods-better alternatives. *Advances in social science methodology* **1**: 43-70.

Hunter D, Yu H, Pukish III M S, Kolbusz J and Wilamowski B M (2012) Selection of proper neural network sizes and architectures—A comparative study. *IEEE Transactions on Industrial Informatics* **8(2)**: 228-240.

Hussain N, Sarfraz S and Javed S (2021) A Systematic Review on Crop-Yield Prediction through Unmanned Aerial Vehicles. *16th International Conference on Emerging Technologies (ICET)*. pp 1-9. IEEE.

Huzsvai L, Zsembeli J, Kovács E and 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)**: 314-333.

Igiri C P, Anyama O U and Silas A I (2015) Effect of learning rate on artificial neural network in machine learning. *International Journal of Engineering Research & Technology* **4(2):** 359-363.

IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Iranmanesh S and Mahdavi M A (2009) A differential adaptive learning rate method for back-propagation neural networks. *World Academy of Science, Engineering and Technology* **50(1):** 285-288.

Jain R C, Agrawal R and Jha M P (1980) Effect of climatic variables on rice yield and its forecast *Mausam* **31(4)**: 591-596.

Jan I, Ashfaq M and Chandio A A (2021) Impacts of climate change on yield of cereal crops in northern climatic region of Pakistan. *Environmental Science and Pollution Research* **28(42)**: 60235-60245.

Jayakumar M and Rajavel M (2017) Coffee yield forecasting using climate indices based agrometeorological model in Kerala. *Mausam* **68(2)**: 309-216.

Jia Q, Li M and 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.

Jia W, Zhao D, Tang Y, Hu C and Zhao Y (2014) An optimized classification algorithm by neural network ensemble based on PLS and OLS. *Mathematical Problems in Engineering* **2014**: 1-8

Johnsson T (1992) A procedure for stepwise regression analysis. *Statistical Papers* **33(1)**: 21-29.

Jolliffe I T (1982) A note on the use of principal components in regression. *Journal of the Royal Statistical Society Series C: Applied Statistics* **31(3)**: 300-303.

Jolliffe I T and Cadima J (2016) Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* **374(2065)**: 20150202.

Jolliffe I T, Trendafilov N T and Uddin M (2003) A modified principal component technique based on the LASSO. *Journal of computational and Graphical Statistics* **12(3)**: 531-547.

Kaiser H F (1960) The application of electronic computers to factor analysis. *Educational and psychological measurement* **20(1)**: 141-151.

Kakati N, Deka R L, Das P, Goswami J, Khanikar P G and Saikia H (2022) Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in the Brahmaputra valley of Assam, North East India. *Theoretical and Applied Climatology* **150(3-4)**: 1201-1215.

Kale V A, Roy S K, Kundu A, Vishnupriya S, Chakraborty M, Hijam L, Maying B, Das S, Rout S, Thapa B and Surje D T (2023) Elucidation of Genetics of Fibre Yield and Quality Traits in Tossa Jute (Corchorus olitorius L.) Using Generation Mean Analysis. *International Journal of Bio-resource and Stress Management* **14(4)**: 530-539.

Kandil N, Khorasani K, Patel R V and Sood V K (1993) Optimum learning rate for backpropagation neural networks. *Proceedings of Canadian Conference on Electrical and Computer Engineering*. pp 465-468. IEEE

Kang Y, Ozdogan M, Zhu X, Ye Z, Hain C and Anderson M (2020) Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environmental Research Letters* **15(6)**: 064005.

Kaul M, Hill R L and Walthall C (2005) Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems* **85(1)**: 1-8.

Kaur B and Gill K K (2017) Development of Weather based Weekly Thumb Rules for Potential Productivity of Mustard Crop in Punjab. *Vayu Mandal* **43(1)**: 72-81.

Kaur S, Diwakar S K and Das A K (2017) Long term rain-fall trend over meteorological subdivisions and districts of India. *Mausam* **68(3)**: 439–450.

Ke J and Liu X (2008) Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction. *Pacific-Asia Workshop on Computational Intelligence and Industrial Application* **2**: 828-832.

Kern A, Barcza Z, Marjanović H, Árendás T, Fodor N, Bónis P, Bognár P and Lichtenberger J (2018) Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agricultural and Forest Meteorology* **260**: 300-320.

Keong Y K and Keng W M (2012) Statistical modeling of weather-based yield forecasting for young mature oil palm. *APCBEE Procedia* **4**: 58-65.

Khan A, Chatterjee S and Bisai D (2017) Air temperature variability and trend analysis by non-parametric test for Kolkata observatory, West Bengal, India. *Indian Journal of Geo Marine Sciences* **46(5)**: 966-971.

Kittichotsatsawat Y, Tippayawong N and Tippayawong K Y (2022) Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. *Scientific Reports* **12(1)**: 14488.

Krishnan A, Williams L J, McIntosh A R and Abdi H (2011) Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56(2)**: 455-475.

Król A (2017) The Application of Partial Least Squares Method in Hedonic Modelling. *Arch Data Sci Ser A* **2**: 1-3

Kross A, Znoj E, Callegari D, Kaur G, Sunohara M, Lapen D R and 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 Sensing* **12(14)**: 2230.

Kumar A and Bhar L (2005) Forecasting model for yield of Indian mustard (Brassica juncea) using weather parameter. *Indian Journal of Agricultural Sciences* **75(10)**: 688-690.

Kumar M, Battan K R and Sheoran O P (2019) Preharvest forecast model for rice yield using principal component regression based on biometrical character with R-software. *International Journal of Agricultural and Statistical Sciences* **15(1)**: 323-326.

Kumar N, Pisal R R, Shukla S P and Pandey K K (2014) Crop yield forecasting of paddy, sugarcane and wheat through linear regression technique for south Gujarat. *Mausam* **65(3)**: 361-364.

Kumar N, Pisal R R, Shukla S P and Pandey K K (2016) Crop yield forecasting of paddy and sugarcane through modified Hendrick and Scholl technique for south Gujarat. *Mausam* **67(2)**: 405-410.

Kumar P, Sahu N C, Kumar S and Ansari M A (2021) Impact of climate change on cereal production: evidence from lower-middle-income countries. *Environmental Science and Pollution Research* **28(37)**: 51597-51611.

Kumar S, Attri S D and Singh K K (2019) Comparison of Lasso and stepwise regression technique for wheat yield prediction. *Journal of Agrometeorology* **21(2)**: 188-192.

Kumar V and Jain S K (2011) Trends in rainfall amount and number of rainy days in river basins of India (1951–2004). *Hydrology Research* **42(4)**: 290-306.

Kumar Y, Singh R, Kumar A and Dhaka A K (2017) Effect of growth and yield parameters on Indian-mustard genotypes under varying environmental conditions in western Haryana. *Journal of Applied and Natural Science* **9(4)**: 2093-2100.

Kunda P, Mukherjee A and Dhal P K (2021) Insights into endophytic bacterial diversity of rice grown across the different agro-ecological regions of West Bengal, India. *World Journal of Microbiology and Biotechnology* **37**: 1-3.

Kundu S K and Mondal T K (1666) Analysis of long-term rainfall trends and change point in West Bengal, India. *Theoretical and Applied Climatology* **138**: 1647-1666.

Kung E C and Sharif T A (1980) Multi-regression forecasting of the Indian summer monsoon with antecedent patterns of the large-scale circulation. *In WMO symposium on probabilistic and statistical methods in weather forecasting* **19(4)**: 295-302.

Kuriqi A, Ali R, Pham Q B, Montenegro Gambini J, Gupta V, Malik A, Linh N T, Joshi Y, Anh D T, Nam V T and Dong X (2020) Seasonality shift and streamflow flow variability trends in central India. *Acta Geophysica* **68**: 1461-1475.

Lacasa J, Messina C D and Ciampitti I (2023) A probabilistic framework for forecasting maize crop yield response to agricultural inputs with sub-seasonal climate predictions. *Environmental Research Letters* **8(7)**: 074042.

Latwal A, Saxena S, Dubey S K, Choudhary K, Sehgal S and Ray S S (2019) Evaluation of pre-harvest production forecasting of mustard crop in major producing states of India, under FASAL project. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **42**: 115-122.

Laxmi R R and Kumar A (2011) Weather based forecasting model for crops yield using neural network approach. *Statistics and Applications* **9(1-2)**: 55-69.

Lecerf R, Ceglar A, López-Lozano R, Van Der Velde M and Baruth B (2019) Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. *Agricultural systems* **168**: 191-202.

Lepcha A, Nag P and Das K K (2020) Can conservation agriculture be the savior of farm economy? a study in North Bengal. *Journal of Crop and Weed* **16(3)**: 123-128.

Lewis-Beck M S (1978) Stepwise regression: A caution. *Political Methodology* **5(2)**: 213-240.

Li C C (1956) The concept of path coefficient and its impact on population genetics. *Biometrics* **12(2)**: 190-210.

Li C C (1968) Fisher, Wright, and path coefficients. *Biometrics* **1**: 471-483.

Li Z, Ding L and Xu D (2022) Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. *Science of The Total Environment* **815**: 152880.

Liao X, Li Q, Yang X, Zhang W and Li W (2008) Multiobjective optimization for crash safety design of vehicles using stepwise regression model. *Structural and multidisciplinary optimization* **35**: 561-569.

Liu B, Zhao Q, Jin Y, Shen J and Li C (2021) Application of combined model of stepwise regression analysis and artificial neural network in data calibration of miniature air quality detector. *Scientific Reports* **11(1)**: 32-47.

Liu C, Zhang X, Nguyen T T, Liu J, Wu T, Lee E and Tu X M (2022) Partial least squares regression and principal component analysis: similarity and differences between two popular variable reduction approaches. *General psychiatry* **35(1)**: e100662.

Lleras C (2005) Path analysis. *Encyclopedia of social measurement* **3(1)**: 25-30.

Lobell D, Cahill K and Field C (2006) Weather-based yield forecasts developed for 12 California crops. *California Agriculture* **60(4)**: 211-215.

Lobell D B and Asseng S (2017) Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters* **212(1)**: 015001.

Lobell D B and Burke M B (2010) On the use of statistical models to predict crop yield responses to climate change. *Agricultural and forest meteorology* **150(11)**: 1443-1452.

Luck J, Asaduzzaman M, Banerjee S, Bhattacharya I, Coughlan K, Chakraborty A, Debnath G C, De Boer R F, Dutta S, Griffiths W and Hossain D (2012) The effects of climate change on potato production and potato late blight in the Asia-Pacific Region. *APN Science Bulletin* **2**: 28-33.

Madeshwaran K, Eswari A, Duraisamy M R, Praneetha S and Selvi B S (2022) Comparison of Linear and Non-linear Models for Coconut Yield Prediction in Coimbatore Using Weather Parameters and External Factors. *International Journal of Environment and Climate Change* **12(11)**: 1141-1150.

Madhu G, Kautish S, Alnowibet K A, Zawbaa H M and Mohamed A W (2023) NIPUNA: A Novel Optimizer Activation Function for Deep Neural Networks. *Axioms* **12(3)**: 246.

Mahesh S, Jayas D S, Paliwal J and White N D (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 and bioprocess technology* **8**: 31-40.

Maitra S and Yan J (2008) Principle component analysis and partial least squares: Two dimension reduction techniques for regression. *Applying Multivariate Statistical Models* **79**: 79-90.

Mall R K, Singh R and Gupta A (2006) Srinivasan G, Rathore LS. Impact of climate change on Indian agriculture: a review. *Climatic change* **78**: 445-478.

Mandal S, Mandal U K, Burman D and Sharma P C (2019) Economic analysis of farm-level agricultural risks in coastal region of West Bengal in India. *Journal of Soil Salinity and Water Quality* **11(2)**: 269-279.

Mansfield E R, Webster J T and Gunst R F (1977) An analytic variable selection technique for principal component regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **26(1)**: 34-40.

Mavromatis T (2016) Spatial resolution effects on crop yield forecasts: An application to rainfed wheat yield in north Greece with CERES-Wheat. *Agricultural Systems* **143**: 38-48.

Meerasri J and Sothornvit R (2022) Artificial neural networks (ANNs) and multiple linear regression (MLR) for prediction of moisture content for coated pineapple cubes. *Case Studies in Thermal Engineering* **33**:101942.

Mehmood T, Saebo S and Liland K H (2020) Comparison of variable selection methods in partial least squares regression. *Journal of Chemometrics* **34(6)**: e3226.

Meinke H and Stone R C (2005) Seasonal and inter-annual climate forecasting: the new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic change* **70(1-2)**: 221-253.

Mevik B H and Wehrens R (2015) Introduction to the pls Package. *Help section of the "Pls" package of R studio software* **22**:1-23.

Mishra V (2013) Food Security Implications of Climate Variability and Climate Change. *Climate Vulnerability: Understanding and Addressing Threats to Essential Resources* **2**: 117-128.

Mitchell J C (1969) The Calculation of Predicted Correlations from Path Coefficients. *Sociology* **3(3)**: 413-416.

Mosteller F and Tukey J W (1977) *Data analysis and regression: A second course in statistics*. Addison-Wesley Pub. Co., Reading, Mass.

Mukherjee A and Huda A K (2018) Assessment of climate variability and trend on wheat productivity in West Bengal, India: crop growth simulation approach. *Climatic change* **147(1-2)**: 235-252.

Mukherjee A, Huda A K and Saha S (2023) Impact of climate change on future productivity and water use efficiency of wheat in eastern India. *Theoretical and Applied Climatology* **152(1-2)**: 421-434.

Murthy C S, Poddar M K, Choudhary K K, Srikanth P, Pandey V, Ramasubramanian S and Kumar G S (2022) Remote sensing based crop insurance for jute (Corchorus olitorius) crop in India. *Remote Sensing Applications: Society and Environment* **26**: 100717.

Nag P, Lepcha A and Das K K (2022) Understanding the Role of Conservation Agriculture with Crop Diversification Strategy in North Bengal Economy. *Insights of Agricultural Sciences* **2022**: 80-95.

Nain G, Bhardwaj N, Jaslam P M and Dagar C S (2021) Rice yield forecasting using agro-meteorological variables: A multivariate approach. *Journal of Agrometeorology* **23(1)**: 100-105.

Nanda A M, Boro Seminar K, Nandika D, Maddu A (2018) A comparison study of kernel functions in the support vector machine and its application for termite detection. *Information* **9(1)**:5.

Ngoc T T, Le Van-Dai Cooch Behar M, Thuyen C M (2021) Support vector regression based on grid search method of hyperparameters for load forecasting. *Acta Polytechnica Hungarica* **18(2)**:143-158.

Niazian M, Sadat-Noori SA, Abdipour M (2018) Modeling the seed yield of Ajowan (Trachyspermum ammi L.) using artificial neural network and multiple linear regression models. *Industrial Crops and Products* **117**: 224-234.

Niles H E (1922) Correlation, causation and Wright's theory of "path coefficients". *Genetics* **7(3)**: 258-273.

Niles H E (1923) The method of path coefficients an answer to wright. *Genetics* **8(3)**: 256-260.

Nishad A, Mishra A N, Chaudhari R, Aryan R K and Katiyar P (2018) Effect of temperature on growth and yield of rice (Oryza sativa L.) cultivars. *International Journal of Chemical Studies* **6(5)**: 1381-1383

Noriega L (2005) *Multilayer perceptron tutorial*. Ph.D. dissertation. School of Computing, Staffordshire University.

Nti I K, Nyarko-Boateng O, Adekoya F A and Weyori B A (2021) An empirical assessment of different kernel functions on the performance of support vector machines. *Bulletin of Electrical Engineering and Informatics* **10(6)**: 3403-3411.

Nwankpa C, Ijomah W, Gachagan A and Marshall S (2021) Activation functions: Comparison of trends in practice and research for deep learning. *2nd International Conference on Computational Sciences and Technology*. pp 124-33. Mehran University of Engineering and Technology Jamshoro, Pakistan.

Oguntunde P G, Lischeid G and Dietrich O (2018) Relationship between rice yield and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis. *International journal of biometeorology* **62(3)**: 459-469.

Olawuwo S, Ogunleye T A, Ojo T O and Adejumo A O (2014) Comparison of Classical Least Squares (CLS), Ridge and Principal Component Methods of Regression Analyses using Gynecological Data. *Journal of Mathematics* **9(6)**: 61-74.

Olusegun A M, Dikko H G and Gulumbe S U (2015) Identifying the limitation of stepwise selection for variable selection in regression analysis. *American Journal of Theoretical and Applied Statistics* **4(5)**: 414-419.

Ozdemir D (2022) The impact of climate change on agricultural productivity in Asian countries: a heterogeneous panel data approach. *Environmental Science and Pollution Research* **1**: 1-13.

Paidipati K K, Chesneau C, Nayana B M, Kumar K R, Polisetty K and 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.

Pal R K, Rao M M and Murty N S (2013) Relative temperature disparity and wheat yield as influenced by sowing environments and genotypes in Tarai Region of Uttarakhand. *Environment and Ecology* **31(2B)**: 979-983.

Pal S and Debnath P (2020) Seasonal incidence of major insect pests and predators of mustard in lower gangetic plains of West Bengal. *Journal of Entomological and Zoological Studies* **8(6)**: 1424-1430.

Pandey K K, Maurya D, Gupta G and Mishra S V (2016) Yield forecasting models based on weather parameters for eastern UP. *Vegetos* **29(1)**: 22-24.

Parviz L (2018) Assessing accuracy of barley yield forecasting with integration of climate variables and support vector regression. *In Annales Universitatis Mariae Curie-Sklodowska, sectio C–Biologia* **73(1)**: 18-30.

Pati S, Saha S, Ghosh A, Pal B, Saha B N and Hazra G C (2021) Silicon availability with two different extractants in intensively rice growing areas of West Bengal, India. *Journal of the Indian Society of Soil Science* **69(3)**: 319-327.

Paudel D, Boogaard H, de Wit A, Janssen S, Osinga S, Pylianidis C and Athanasiadis I N (2021) Machine learning for large-scale crop yield forecasting. *Agricultural Systems* **187**: 103016.

Paul P and Paul S (2020) Level of Human Development in North Bengal: A Geographical Appraisal. *Geography and Development* : 128-139

Pavithra S and Vengadessan V (2020) Selection for drought tolerance in rice genotypes based on principal components and seletion indices. *Electronic Journal of Plant Breeding* **11(4)**: 1032-1036.

Pence I, Kumaş K, Siseci M C and Akyüz A (2022) Modeling of energy and emissions from animal manure using machine learning methods: the case of the Western Mediterranean Region, Turkey. *Environmental Science and Pollution Research* **27**: 1-22.

Poddar S, Nath S, Mandal R, Roy S K and Das S (2022) Genetic Component Analysis and Determination of Optimum Number of Clusters Based on Morpho-Physiological Traits in Wheat. *International Journal of Bio-resource and Stress Management* **13(12)**: 1440-1449.

Poonam A D, Khichar M L and Mamta T N (2017) Statistical models for wheat yield forecasting using weather parameters for western agroclimatic zone of Haryana. *Bulletin of Environment, Pharmacology and Life Sciences* **6(4)**: 424-427.

Pope P T and Webster J T (1972) The use of an F-statistic in stepwise regression procedures. *Technometrics* **14(2)**: 327-340.

Pradel W, Gatto M, Hareau G, Pandey S K and Bhardway V (2019) Adoption of potato varieties and their role for climate change adaptation in India. C*limate Risk Management* **23**: 114-123.

Puig-Arnavat M and Bruno J C (2015) Artificial neural networks for thermochemical conversion of biomass. *Recent advances in thermo-chemical conversion of biomass* **2015**: 133-156.

Qu H and Zhang Y (2016) A new kernel of support vector regression for forecasting high-frequency stock returns. *Mathematical Problems in Engineering* **:** 1-9.

Radhakrishnan K, Sivaraman I, Jena S K, Sarkar S and Adhikari S (2017) A climate trend analysis of temperature and rainfall in India. *Climate Change and Environmental Sustainability* **5(2):** 146–153.

Raha G N, Bhattacharjee K, Das M, Dutta M and Bandyopadhyay S (2014) Statistical study of surface temperature and rainfall over four stations in north Bengal. *Mausam* **65(2)**: 179-184.

Rahman S and Khan M R (2012) Incidence of pests in jute (Corchorus olitorius L.) ecosystem and pest–weather relationships in West Bengal, India. *Archives of Phytopathology and Plant Protection* **45(5)**: 591-607.

Rai S, Nandre J and Kanawade B R (2022) A Comparative Analysis of Crop Yield Prediction using Regression. *2nd International Conference on Intelligent Technologies (CONIT).* pp 1-4. KLE Institute of Technology, Hubli, Karnataka, India.

Rajavel M, Khare P, Prasad J R, Singh K K, Puranik H V and Das G K (2018) Development of rice yield forecast in mid-season using weather indices based agrometeorological model in Chhattisgarh. *Vayu Mandal* **44(1)**: 38-45.

Rea A and Rea W (2016) How many components should be retained from a multivariate time series PCA?. *arXiv preprint arXiv:1610.03588*.

Roberts M J, Braun N O, Sinclair T R, Lobell D B and Schlenker W (2017) Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters* **12(9)**: 095010.

Robinson R (1991) Neural networks offer an alternative to traditional regression. *Geobyte* **7**: 14-19.

Rotter R P, Carter T R, Olesen J E and Porter J R (2011) Crop–climate models need an overhaul. *Nature climate change* **1(4)**: 175-177.

Roy B (2019) Characterization of farmers' varieties of potato (Solanum tuberosum L.) of Cooch Behar district of West Bengal. *Indian Journal of Plant Genetic Resources* **32(3)**: 314-317.

Roy S K, Chakraborty M, Hijam L, Mondal H A, Surje D T, Roy A, Mondal R, Pal S, Kundu A, Das S and Sarkar P (2018) Studies on genetic variability and screening for fibre yield components and biotic stress factors in tossa jute (Corchorus olitorius L.) germplasm under Terai region of West Bengal. *Electronic Journal of Plant Breeding* **9(2)**: 409-423.

Ruengvirayudh P and Brooks G P (2016) Comparing stepwise regression models to the best-subsets models, or, the art of stepwise. *General Linear Model Journal* **42(1)**: 1-14.

Rumelhart D E (1986) Learning internal representations by error propagation, in parallel distributed processing: Explorations in the Microstructure of Cognition. pp 318-362. MIT Press.

Sagar B M and Cauvery N K (2018) Agriculture data analytics in crop yield estimation: a critical review. *Indonesian Journal of Electrical Engineering and Computer Science* **12(3)**: 1087-1093.

Saha A, Singh K N, Ray M, Rathod S and Choudhury S (2021) Modelling and forecasting cotton production using tuned-support vector regression. *Current Science* **121(8)**: 1090-1098

Sam K and Chakma N (2019) An exposition into the changing climate of Bengal Duars through the analysis of more than 100 years' trend and climatic oscillations. *Journal of Earth System Science* **128**: 1-2.

Samui P, Sitharam T G and Kurup P U (2008) OCR prediction using support vector machine based on piezocone data. *Journal of Geotechnical and Geo Environmental engineering* **134(6)**: 894-898.

Sánchez B, Rasmussen A and Porter J R (2014) Temperatures and the growth and development of maize and rice: a review. *Global change biology* **20(2)**: 408-417.

Satish M S, Bhambhar K S, Yadavrao W S, Shaikh A A, Balasubramanian R and Dadaji J J (2023) Assessment of Crop Weather Relations in Wheat (Triticum aestivum) in Western Maharashtra Plain Zone. *International Journal of Environment and Climate Change* **13(8)**: 502-510.

Satpathi A, Setiya P, Das B, Nain A S, Jha P K, Singh S and Singh S (2023) Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India. *Sustainability* **15(3)**: 2786.

Savas C and Dovis F (2019) The impact of different kernel functions on the performance of scintillation detection based on support vector machines. *Sensors* **19(23)**: 5219.

Sclove S L (2022) Determining an Adequate Number of Principal Components. *Advances in Principal Component Analysis* **2022**:1-15.

Semenov M A and Porter J R (1995) Climatic variability and the modelling of crop yields. *Agricultural and forest meteorology* **73(3-4)**: 265-283.

Setiya P, Satpathi A, Nain A and Das B (2022) Comparison of weather-based wheat yield forecasting models for different districts of Uttarakhand using statistical and machine learning techniques. *Journal of Agrometeorology* **24(3)**: 255-261.

Shafiee S, Lied L M, Burud I, Dieseth J A, Alsheikh M and Lillemo M (2021) Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Computers and Electronics in Agriculture* **183**: 106036.

Shahhosseini M, Hu G and Archontoulis S V (2020) Forecasting corn yield with machine learning ensembles. *Frontiers in Plant Science* **11**: 1120.

Shahsavari M, Mohammadi V, Alizadeh B and Alizadeh H (2023) Application of machine learning algorithms and feature selection in rapeseed (Brassica napus L.) breeding for seed yield. *Plant Methods* **19(1)**: 57.

Sharabiani S V, Soltani Nazarloo A, Taghinezhad E, Veza I, Szumny A and Figiel A (2023) Prediction of winter wheat leaf chlorophyll content based on VIS/NIR spectroscopy using ANN and PLSR. *Food Science & Nutrition* **11(5)**: 2166-2175.

Sharma P, Singh B K and Singh R P (2018) Prediction of potato late blight disease based upon weather parameters using artificial neural network approach. *9th International Conference on Computing, Communication and Networking Technologies* **2018**: 1-13.

Sharma S, Sharma S and Athaiya A (2017) Activation functions in neural networks. *Towards Data Science* **6(12)**: 310-316.

Sharma S K, Bhagat D V, Ranjeet P D, Khapedia H L, Mirdha I S and Sikarwar R S (2018) Soybean and wheat crop yield forecasting based on statistical model in Malwa agroclimatic zone. *International Journal of Chemical Studies* **6(4)**: 1070-1073.

Sheela K G and Deepa S N (2013) Review on methods to fix number of hidden neurons in neural networks. *Mathematical problems in engineering* **2013**: 1-11.

Shetty N and Gislum R (2011) Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR. *Field Crops Research* **120(1)**: 31-37.

Siebert S, Webber H and Rezaei E E (2017) Weather impacts on crop yields-searching for simple answers to a complex problem. *Environment Research Letter* **12(8)**: 081001.

Silaban H and Zarlis M (2017) Analysis of Accuracy and Epoch on Back-propagation BFGS Quasi-Newton. *In Journal of Physics: Conference Series* **930(1)**: 012006.

Singh A, Nagar V, Tanwar N, Shahi U P and Bhan S C (2021) Use of Statistical Models in Yield Forecasting of Wheat, Mustard and Potato Crop in Western Districts of Uttar Pradesh, India. *International Journal of Environment and Climate Change* **11(12)**: 573-579.

Singh A K, Barman D, Behera M S, Mazumdar S P, Saha A R and Kundu D K (2018) Impact of climate change on productivity of tropical rice-wheat-jute system under long term fertilizer management in alluvial soils. *International Journal of Current Microbiology and Applied Sciences* **7(11)**: 1623-1632.

Singh K N, Singh K K, Sudheer K, Sanjeev P and Bishal G (2019) Forecasting crop yield through weather indices through LASSO. *Indian Journal of Agricultural Sciences* **89(3)**: 540-544.

Singh D and Sandhu S K (2022) Effect of Climate Variability on Potato Yield Under Different Agroclimatic Zones of Punjab. *Potato Journal* **49(2)**: 159-166.

Singh S V, Marwaha R S, Kumar D, Kumar P and Pandey S K (2009) Suitability of potato varieties grown in north-eastern Indian plains for processing. *Potato Journal* **36(1-2)**: 25-34.

Smith G (2018) Step away from stepwise. *Journal of Big Data* **5(1)**: 1-12

Smola A J and Schölkopf B (2004) A tutorial on support vector regression. *Statistics and computing* **14**: 199-222.

Sridhara S, Manoj K N, Gopakkali P, Kashyap G R, Das B, Singh K K and Srivastava A K (2023) Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. *International Journal of Biometeorology* **67(1)**: 165-180.

Sridhara S, Ramesh N, Gopakkali P, Das B, Venkatappa S D, Sanjivaiah S H, Singh K, Singh P, El-Ansary D O, Mahmoud E A and Elansary H O (2020) Weather-based neural network, stepwise linear and sparse regression approach for rabi sorghum yield forecasting of Karnataka, India. *Agronomy* **10(11)**: 1645-1670.

Srinivasan P A, Guastoni L, Azizpour H, Schlatter P H and Vinuesa R (2019) Predictions of turbulent shear flows using deep neural networks. *Physical Review Fluids* **4(5)**: 054603.

Struik P C (2007) Responses of the potato plant to temperature. *Potato Biology and Biotechnology: Advances and Perspectives*. pp 367-393. Elsevier Science BV.

Sujaritha M, Kavitha M and Janet J (2019) Comparative Analysis of Accuracy on Partial Least Squares and Principal Component Regression methods. *Indian Journal of Science and Technology* **12(8)**: 1-9.

Sur B, Rout S, Singla S, Mandal R, Nath S, Maying B, Sadhu S, Chakraborty M, Hijam L, Debnath M K and Roy S K (2023) Evaluation of mustard genotypes [Brassica juncea (L.) Czern and Coss] for quantitative traits and character association of seed yield and yield components at sub Himalayan region of West Bengal (India). *Plant Science Today* **10(1)**: 166-178.

Surendran U, Raja P, Jayakumar M and Subramoniam S R (2021) Use of efficient water saving techniques for production of rice in India under climate change scenario: A critical review. *Journal of Cleaner Production* **309**: 127272.

Szandała T (2021) Review and comparison of commonly used activation functions for deep neural networks. *Bio-inspired neurocomputing* : 203-224.

Tan B T, Fam P S, Firdaus R R, Tan M L and Gunaratne M S (2021) Impact of climate change on rice yield in Malaysia: a panel data analysis. *Agriculture* **11(6)**: 569-576.

Taşan S, Cemek B, Taşan M and Cantürk A (2022) Estimation of eggplant yield with machine learning methods using spectral vegetation indices. *Computers and electronics in agriculture* **202**: 107367.

Trnka M, Olesen J E, Kersebaum K C, Rötter R P, Brázdil R, Eitzinger J, Jansen S, Skjelvåg A O, Peltonen-Sainio P, Hlavinka P and Balek J (2016) Changing regional weather crop yield relationships across Europe between 1901 and 2012. *Climate Research* **70(2-3)**: 195-214.

Truong V H and Pham H A (2021) Support vector machine for regression of ultimate strength of trusses: A comparative study. *Engineering Journal* **25(7)**: 157-166.

Turner M E and Stevens CD (1959) The regression analysis of causal paths. *Biometrics* **15(2)**: 236-258.

Valle S, Li W and Qin S J (1999) Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods. *Industrial & Engineering Chemistry Research* **38(11)**: 4389-4401.

Vander-Maaten L, Postma E and Van den Herik J (2009) Dimensionality reduction: a comparative. *Journal of Machine Learning Research* **10(66-71)**: 1-35.

Vander-Velde M, Biavetti I, El-Aydam M, Niemeyer S, Santini F and van den Berg M (2019) Use and relevance of European Union crop monitoring and yield forecasts. *Agricultural systems* **168**: 224-230.

Van-Klompenburg T, Kassahun A and Catal C (2020) Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* **177**: 105709.

Vapnik V (2000) SVM method of estimating density, conditional probability, and conditional density. *In 2000 IEEE International Symposium on Circuits and Systems (ISCAS)* **2**: 749-752.

Vashisth A (2021) Mustard yield prediction using machine learning approach. *Journal of Agricultural Physics* **21(2)**: 445-456.

Vashisth A and Goyal A (2023) Prediction of mustard yield using different machine learning techniques: a case study of Rajasthan, India. *International Journal of Biometeorology* **67(3)**: 539-551.

Vatcheva K P, Lee M, McCormick J B and Rahbar M H (2016) Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)* **6(2)**: 227-246.

Verma U, Ruhal D S, Hooda R S, Yadav M, Khera A P, Singh C P, Kalubarme M H and Hooda I S (2003) Wheat yield modelling using remote sensing and agrometeorological data in Haryana State. *Journal of Indian Society of Agricultural Statistics* **56(2)**: 190-198.

Vora K, Yagnik S and Scholar M (2014) A survey on backpropagation algorithms for feedforward neural networks. *International Journal of Engineering Development and Research* **1(3)**: 193-197.

Wakeling I N and Morris J J (1993) A test of significance for partial least squares regression. *Journal of Chemometrics* **7(4)**: 291-304.

Wang F, Shao W, Yu H, Kan G, He X, Zhang D, Ren M and Wang G (2020) Re-evaluation of the power of the mann-kendall test for detecting monotonic trends in hydrometeorological time series. *Frontiers in Earth Science* **8**: 1-12.

Wang P, Chen S and Yang S (2022) Recent advances on penalized regression models for biological data. *Mathematics* **10(19)**: 3695.

Wang W and Li H (2020) Partial Least Squares Regression for Predicting the Speed of Electromagnetic Fuze Plate. *Journal of Physics: Conference Series* **1549(3)**: 032084.

Wang X, Folberth C, Skalsky R, Wang S, Chen B, Liu Y, Chen J and Balkovic J (2022) Crop calendar optimization for climate change adaptation in rice-based multiple cropping systems of India and Bangladesh. *Agricultural and Forest Meteorology* **315**: 108830.

Wentzell P D and Montoto L V (2003) Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemometrics and intelligent laboratory systems* **65(2)**: 257-279.

Wheeler R M, Tibbitts T W and Fitzpatrick A H (1989) Potato growth in response to relative humidity. *HortScience* **24(3)**: 482-484.

Wickramasinghe L, Weliwatta R, Ekanayake P and Jayasinghe J (2021) Modeling the relationship between rice yield and climate variables using statistical and machine learning techniques. *Journal of Mathematics* **2021**: 1-9.

Wold S, Esbensen K and Geladi P (1987) Principal component analysis. *Chemometrics and intelligent laboratory systems* **2(1-3)**: 37-52.

Wood M D, Simmatis L E, Jacobson J A, Dukelow S P, Boyd J G and Scott S H (2021) Principal components analysis using data collected from healthy individuals on two robotic assessment platforms yields similar behavioral patterns. *Frontiers in human neuroscience* **15**: 652201.

Wright S (1923) Correlation and causation. *Journal Agricultural Research* **7**: 557-585.

Wright S (1934) The method of path coefficients. *The annals of mathematical statistics* **5(3)**: 161-215.

Wright S (1960) Path coefficients and path regressions: alternative or complementary concepts?. *Biometrics* **16(2)**: 189-202.

Yadav R R, Sisodia B V and Kumar S (2014) Application of principal component analysis in developing statistical models to forecast crop yield using weather variables. *Mausam* **65(3)**: 357-360.

Yang K, Li J and Wang C (2006) Missing values estimation in microarray data with partial least squares regression. *In Computational Science–ICCS 2006: 6th International Conference*. **3992**, pp 662–669. Springer, Berlin, Heidelberg.

Yekkehkhany B, Homayouni S, McNairn H and Safari A (2014) Multi-temporal full polarimetry L-band SAR data classification for agriculture land cover mapping. *Geoscience and Remote Sensing Symposium*. pp 2770-2773. IEEE.

Yeniay Ö and Göktaş A (2002) A comparison of partial least squares regression with other prediction methods. *Hacettepe Journal of Mathematics and Statistics* **31**: 99-111.

Yue S and Wang C (2004) The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water resources management* **18(3)**: 201-218.

Zhang Z (2014) Too much covariates in a multivariable model may cause the problem of overfitting. *Journal of thoracic disease*. **6(9)**: E196.

Zifarelli A, Giglio M, Menduni G, Sampaolo A, Patimisco P, Passaro V M, Wu H, Dong L and Spagnolo V (2020) Partial least-squares regression as a tool to retrieve gas concentrations in mixtures detected using quartz-enhanced photoacoustic spectroscopy. *Analytical Chemistry* **92(16)**: 11035-11043.

Zou H and Hastie T (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **67(2)**: 301-320.

****

Document Information

Analyzed document	Thesis of Ajith S (Reg. No A-2020-012-D).pdf (D172169933)
Submitted	2023-07-17 10:57:00
Submitted by	MANOJ KANTI DEBNATH
Submitter email	mkanti1984@gmail.com
Similarity	1%
Analysis address	mkanti1984.ubkv@analysis.urkund.com

Sources included in the report

SA	**Manuscript.docx** Document Manuscript.docx (D157976581)	1
SA	**Thesis by Manoj Varma_21238.docx** Document Thesis by Manoj Varma_21238.docx (D112965534)	5
SA	**Merlin-thesis.pdf** Document Merlin-thesis.pdf (D142876572)	5
SA	**updated manuscript2.docx** Document updated manuscript2.docx (D129728625)	9
SA	**Manuscript_P.Das_21-06-2022.docx** Document Manuscript_P.Das_21-06-2022.docx (D140902845)	3
SA	**FORECASTING MODELS-GROUNDNUT.docx** Document FORECASTING MODELS-GROUNDNUT.docx (D163633923)	1
SA	**Maize-research paper 25.3.2022.docx** Document Maize-research paper 25.3.2022.docx (D131495068)	2
SA	**Revised Research paper 02.06.2022.docx** Document Revised Research paper 02.06.2022.docx (D138949379)	2
SA	**All chapters word.docx** Document All chapters word.docx (D127928730)	1
SA	**Vijay_Dept of Farm engineering,IAS.docx** Document Vijay_Dept of Farm engineering,IAS.docx (D44500019)	2
SA	**research paper (2).docx** Document research paper (2).docx (D94907841)	2

VITA

Name of the student	:	AJITH S
Father's name	:	SAKTHIVEL C
Mother's name	:	SUSILA S
Nationality	:	INDIAN
Date of birth	:	09-05-1996
Permanent home address	:	2/163, Mel Street, Kadagathur (P.O.), Dharmapuri (Tk. & Dt.), Tamil Nadu-636809.

EDUCATION QUALIFICATIONS

Bachelor degree	:	B.Sc. (Ag.)
Name of the university	:	Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu.
Year of completion	:	2017
OGPA/CGPA/ % of Marks	:	8.37/10.00
Master degree	:	M.Sc. (Ag.) in Agricultural Statistics
Name of the university	:	Acharya N.G. Ranga Agricultural University, Guntur, Andhra Pradesh.
Year of completion	:	2019
OGPA/CGPA/ % of Marks	:	9.11/10.00
Title of Master's thesis	:	EFFECT OF NATIONAL FOOD SECURITY MISSION ON FOOD GRAIN PRODUCTION IN INDIA
Doctor of Philosophy	:	Ph.D. (Ag.) in Agricultural Statistics
OGPA	:	9.09/10.00