Plant Archives Vol. 24, Special Issue (GABELS), 2024 pp.461-467 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org

DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2024.v24.SP-GABELS.066

DEVELOPMENT OF GROUNDNUT YIELD PREDICTING MODEL IN RELATION TO WEATHER PARAMETERS IN DHARWAD DISTRICT, KARNATAKA INDIA

Ashwini. C.B.*, Ashalatha K.V., and Megha J.

Department of Agricultural Statistics, University of Agricultural Sciences, Dharwad, Karnataka India

*Corresponding author Email: ashalathakv@uasd.in

ABSTRACT

Groundnut (*Arachis hypogea* L.) is a high-yielding cash crop with multiple uses for each plant part, including direct consumption, cooking oil and a rich source of protein feed for livestock. Keeping that in view, six models were developed for two seasons of Dharwad districts of Karnataka to predicting the groundnut yield viz., Linear Regression, Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ELNET) regression, Support Vector Regression and K-Nearest Neighbor. In this study an attempt has been made to study the relationship for weather parameterson groundnut yield (*kharif* and *summer)* for more than four decades. The assessment of models was done by fixing 80% of the data for calibration and left 20 % data for validation. Specifically, the study highlighted the significantly positive impact of relative humidity and rainfall on kharif season yields, along with the significantly negative influence of temperature whereas in summer season significant negative impact on summer yield. Moreover, it identified LASSO regression as a promising model for predicting summer groundnut yields, showcasing its potential in optimizing crop management. Similarly, for kharif groundnut yield prediction K-NN turned to be the best model.

Keywords **:** Linear Regression, Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ELNET) regression, Support Vector Regression and K-Nearest Neighbor.

Introduction

Statistical analysis entails a structured procedure for gathering and examining data, aimed at revealing patterns and trends. By modelling current circumstances and projecting potential future scenarios, this practice aids decision-making for businesses, policymakers and authorities. It assists in identifying optimal performance, refining existing products, eliminating inefficiencies and innovating new products. The enigmatic nature of statistical data is particularly crucial in agricultural and related domains. Effectively addressing the diverse challenges in various facets of agriculture demands the application of statistical methods and approaches.

Yield, a complex trait, is influenced by numerous contributing factors and their interactions. Environmental conditions, along with a set of interconnected traits regulated by a few genes, have a considerable impact on yield. Studying correlations among traits aids in selecting multiple traits simultaneously. The spectra of climate change pose a significant threat to long-term development, with the agricultural sector being particularly vulnerable (Reyer *et al.*, 2017). The sector's susceptibility has led to widespread poverty and food insecurity among rural populations. Tackling climate change requires comprehension of its causes and consequences to devise adaptation and mitigation strategies, fostering sustainable agricultural production and addressing hunger (Raza *et al.*, 2019).

Beyond air temperature, crop yields are significantly influenced by soil water content and its storage capacity (Lipiec *et al.*, 2013). Factors like rain absorption by soil, evaporation from the soil surface and water inaccessible to crops shape yield outcomes. Air humidity affects crop transpiration; drier air curtails dry matter production. Variations in rainfall and humidity wield substantial influence over crop yields. Climate change extends beyond average temperature and rainfall; it also alters the frequency of extreme weather events, which can inflict significant costs and damage on agriculture. Profitability in commercial farming often hinges on the balance between favorable and adverse weather conditions.

Agriculture is one of the greatest susceptible sectors to climatic change as weather variables involving temperature and rainfall are direct inputs into agriculture production system (Saravanakumar et al., 2022). Groundnut (*Arachis hypogaea* L.) is a highly productive cash crop with versatile applications for each of its plant parts, encompassing direct consumption, cooking oil and a valuable protein source for livestock feed. India is the second-largest global producer of groundnuts after China, followed by the United States and Nigeria. In the Indian context, groundnut holds significance as a primary rain-fed protein crop. According to Kalarani *et al*. (2023), India cultivates groundnuts during the summer (25%) of production and Kharif (75%) seasons, yielding two harvests in March and October. Karnataka, a significant groundnut-producing state in India, contributes around 1.4 million tons annually, accounting for approximately 18 percent of the nation's total groundnut production. Key groundnut-growing districts in Karnataka include Dharwad, Haveri, Belgaum and Chitradurga. For Karnataka's farmers, especially those in rainfed regions, groundnut holds considerable importance. The crop spans approximately 0.74 million hectares, yielding 0.72 million Tons and achieving a productivity of 979.9 kg/ha (DES, 2021).

The fluctuations in theweather variables may exert adverse impact on the groundnut yields (Ravi *et al*., 2023). This calls for prediction of groundnut yields in tune with weather variability in different season of Dharwad district Karnataka. Although, several researchers have developed statistical and machine learning models for various crops including groundnut there is a need to compare the performance of well-developed conventional statistical methods and ML approaches in the domain of crop yield due to the fact that the performance of statistical methods differs regionally (spatially). Machine learning is a part of artificial intelligence employed to build an intelligent system (Chlingaryan *et al*., 2018). Hence, it is important to study the accuracy of different models for predicting groundnut separately for study area (Sridhara *et al*., 2023) by using six different ML models. Accordingly, different yield forecasting models have been employed in Dharwad district region to enable the farmers and other stakeholders to employ the best-fitted model to utilize weather information in a timely manner for successful planning and groundnut farming decision-making.

Material and Methods

Dharwd is located in the North-Western part of Karnataka state with coordinates 15˚27'30" North and 75˚00'30" East. Dharwad experiences tropical wet climate, which is 741mabove sea level. The south-west monsoon is most crucial for Dharwad district. Average annual rainfall of the district was 864 mm. The average temperature is 24.1˚C. The climate is mildly hot during the summer (April–May) and pleasant during rest of the year. The hot season begins from during the month of December, which is the coldest month.

The present study was based on the secondary data on weather parameters Temperature (°C), Relative humidity (%), Rainfall (mm), Wind speed (km/hr), and latent heat flux (W/m^2)) over a period of 41 years (1980-2021) which was collected Weather parameters data were collected from the Department of Agrometeorology, UAS, Dharwad, CMIP (Climate Model Intercomparison Project). The *kharif* and *summer* groundnut yield data for the period of 41 years (1980-2021) were collected from District Statistical office of Dharwad district of Karnataka.

Correlation analysis

Correlation measures the degree of closeness or association between two variables and the strength of the relationship between different parameters. Selected variable to meet their assumptions, correlation heatmaps are generated in the form of a matrix with radiant colours, if darker colour signifies stronger correlation, then lighter colour signifies weak correlation and vice-versa. The correlation heatmap was generated using seaborn library of python language.

Linear Regression

The linear regression models were fitted using crop yield as dependent variable (Y) and weather parameters as independent variables (X_1, X_2,..,X_5). The linear regression equation is given by, (Nageshwara Rao,2008)

$$Y_i = \alpha + \beta X + \varepsilon \quad (1.1)$$

Where, Y isthe dependent variable, α intercept parameter, X is independent variable with β as the simple regression coefficient of Yon X and ε is the error term.

LASSO regression

It is a method that combines the least-squares loss with an L_1- constraint, or bound on the sum of the absolute values of the coefficients. Relative to the least-squares solution, this constraint has the effect of shrinking the coefficients and even setting some to zero. In this way it provides an automatic way for doing model selection in linear regression. Moreover, unlike some of other criteria for the model selection, the resulting optimization problem is convex and can be solved efficiently for large problems. Given a collection of N predictor-response pairs $\{(x_i, y_i)\}$ from i = 1 approaching to N, the lasso finds the solution (β_0, β_j) to the optimization problem.

$$minimize \left\{ \frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \right\} \quad (1.2)$$

$$subject\ to \sum\nolimits_{j=1}^{p} |\beta_j| \le t$$

Ridge regression

Ridge regression causes the regression coefficients to shrink so that factors that have a negligible impact on the outcome have their coefficients near to zero. The reduction of the coefficients is accomplished by penalizing the regression model with a term known as L_2-norm, which is the sum of the squared coefficients (Zou and Hastie, 2005). Where, the loss is defined as:

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2 = y - X\hat{\beta}^2 + \lambda\hat{\beta}^2 \quad (1.3)$$

where represents the independent variable, β represents the coefficient associated with it, and λ represents the L^2 norm penalty.

Elastic net (ELNET) regression

The ELNET model has features of both LASSO and ridge regressions i.e., it considers both the L1 and L2 norms. (Hoerl and Kennard, 1970). This causes some coefficients to shrink and some coefficients to be set to zero. Therefore, it reduces the impact of various features without eliminating them completely (Cho *et al.*, 2009).

$$L_2 = \sum \left(\hat{Y}_i - Y_i \right)^2 + \lambda \sum \beta^2 \quad (1.4)$$

where, represents the independent variable, β represents the corresponding coefficient and λ represents the penalty

Support Vector Regression

Support Vector Regression introduced by Vladimir N. Vapnik and Alexey Ya. Chervonenk is in 1963, is the regression model of Support Vector Machine, on a dataset consisting of L samples of form $\{(x_1, y_2), (x_1, y_2), \ldots, (x_L, y_L), x \in R^m, y \in R\}$ is a linear function which can estimate output values based on inputs.

$$y = (w.x) + b \quad (1.5)$$

where, y is the estimated value, x is the input vector, w is the weight vector and b are the bias.

SVR creates a hyperplane or set of hyperplanes in a high or infinite dimensional space, which is utilized for regression, classification or other tasks. SVR uses linear functions for learning. In case of nonlinear cases, SVR uses a kernel technique to plot the data into a higher dimensional feature space, in which linear functions can be applied (Palanivel *et al*., 2019).

While using the SVM for regression analysis, a margin of tolerance is fixed in approximation to the SVM which has been used for the problem. SVMs represent the hyperplane as optimized hyperplanes with support vectors (Mohapatra and Chaudhary, 2022).

Hyper parameters in SVR:

1. Hyperplane: Hyperplanes are decision boundaries for predicting the continuous output. Support Vectors are the data points on either side of the hyperplane that are closest to the hyperplane. These are used to draw the required line that shows the algorithm's predicted outcome.

2. Kernel: A kernel is a collection of mathematical functions that take data and changeit into the desired form. These are most commonly used to find a hyperplane in higher-dimensional space. Linear, Non-Linear, Polynomial, Radial Basis Function (RBF) and Sigmoid are the most commonly used kernels. RBF is the kernel that is used by default. Depending on the dataset, each of these kernels are used.

3. Boundary Lines: These are the two lines that are drawn at a distance of ε (epsilon) from the hyperplane. It's used to separate the data points by a margin.

4. Support Vectors: The closest point of the lines from both the classes.

The Support Vector Regressor model fits the hyperplane which has the maximum points and uses a threshold value. The Support Vector Machine algorithm model can also be useful as a regression analysis technique while considering the significant features that can characterize the algorithm.

K-Nearest Neighbor

Cover & Hart introduced K-Nearest Neighbour which is a machine learning method used for regression as well as classification. K-NN considers each data record as a vector in an m-dimensional space (where m is the number of features) and predicts the value of each new sample based on the values of K records that are closest to that point in that space (Enas and Choi, 1986).

The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the K closest data points to the new observation, and to take the most common class among these (Karthikeya*et al*., 2020). This is why it is called the K-Nearest Neighbors algorithm. This distance is calculated using various measures such as Euclidean distance, Minkowski distance, Mahalanobis distance. The larger is K; the better is classification. For instance, the closeness of the new point x and the training point x_i is measured by a Euclidean distance function in the form of equation as following

$$d\,(x, x_i) = \sqrt{\sum_{j=1}^{m}(x_i^j - x^j)^2} \qquad ....(1.6)$$

i=1, 2,.,n.and j=1,2,..,m.

Where n is the number of training samples and m is the number of input samples. Samples that are closer to the new sample will have a greater impact on the prediction. The implementation of algorithm can be noted as below:

1. Load the data.
2. Initialize K to your chosen number of neighbors.
3. For each example in the data.
4. Calculate the distance between the query example and the current example from the data.
5. Add the distance and the index of to an ordered collection.
6. Sort the ordered collection of distances and indices from smallest to largest (in ascending order)by the distances.
7. Pick the first K entries from the sorted collection.
8. Get the labels of the selected K entries.
9. If regression, return the mean of the K label.

Results and Discussion

The nature and degree of association of weather parameter with yield of different season crop and among different weather parameters are studied. The degree of relationship between the weather variables and production of crop in kharif season of Dharwad district is presented in the form of correlation matrix heatmap (Fig.1a).

There was a strong association observed among the weather variables, showing a possible presence of multicollinearity, which could hinder the results of regression analysis further. For kharif crop, relative humidity and rainfall were significant and positively correlated with the production indicating that increase in amount of rainfall and relative humidity could result in significant rise in yield of the crop and temperature showed highly negative significant that is rise in temperature decrease in yield of the groundnut (Table 1).

As presented in the form of correlation heatmap (Fig.1b) for summer crop, an attempt on correlating the weather parameters was made with the yield of crop (Table 2). There was again strong relationship observed among weather variables for the crop. For summer crop, again temperature was significant and negatively correlated with the yield signifying that increase in temperature could result in drop in production of crop. These results were on par with study conducted by Pavan (2022), revealed than significant and very high positive correlation was observed between rainfall and relative humidity for crop yield, while in some district, non-significant positive correlation was observed.

Table 1: Correlation of weather parameters and kharif season production of groundnut in Dharwad district

	Yield	Temperature	Rainfall	Relative Humidity	Wind Speed	Latent Heat Flux
Yield	1					
Temperature	-0.411**	1				
Rainfall	0.359*	0.153	1			
Relative Humidity	0.385*	0.141	0.729**	1		
Wind Speed	-0.138	-0.294	-0.352*	-0.416**	1	
Latent Heat Flux	0.212	-0.730**	-0.196	-0.181	0.124	1

Note: ** significant at 1% level * significant at 5% level

Table 2: Correlation of weather parameters and summer season production of groundnut in Dharwad district

	Yield	Temperature	Rainfall	Relative Humidity	Wind Speed	Latent Heat Flux
Yield	1					
Temperature	-0.354*	1				
Rainfall	0.099	-0.353*	1			
Relative Humidity	0.256	-0.367*	0.463**	1		
Wind Speed	0.057	-0.166	0.197	0.176	1	
Latent Heat Flux	0.156	-0.446**	0.240	0.133	0.498**	1

Note: ** significant at 1% level, * significant at 5% level

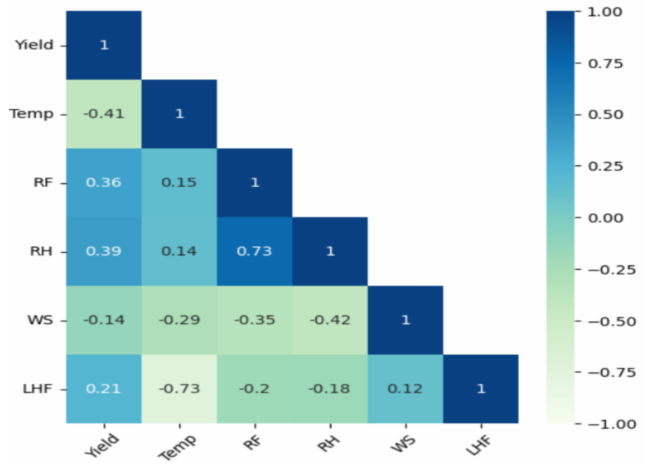

Fig. 1(a): Correlation heatmaps of weather parameters and production of kharif groundnut in Dharwad district

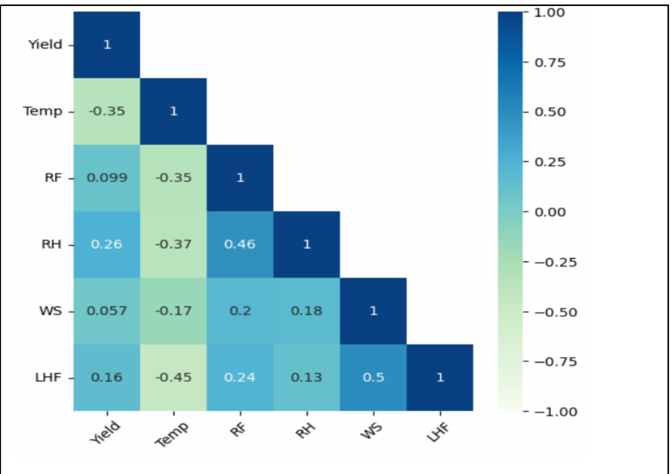

Fig. 1(b): Correlation heatmaps of weather parameters an production of summer groundnut in Dharwad district

Different machine learning algorithms were evaluated for prediction of groundnut yield based on weather parameters and best model was selected based on the values obtained in the form of evaluation metrics.The dataset was divided into 80:20 ratio of training and testing dataset and testing dataset was used for evaluation of the dataset. The model which has the lowest values of MSE, RMSE, MAE and MeAE is considered the best model for prediction of groundnut yield based on weather parameters.

Table 3: Comparison of different machine learning algorithms for prediction of groundnut yield in Kharif season.

Models	MSE	RMSE	MAE	MeAE
Linear regression	2001.206	44.735	34.766	22.313
LASSO	2000.568	44.728	34.759	22.297
Ridge Regression	1214.459	34.849	23.495	18.701
ELNET	1642.771	40.530	30.193	18.318
SVR	1453.780	38.128	25.920	12.596
KNN	**1127.238**	**33.574**	**22.766**	**12.194**

Least values of MSE, RMSE, MAE and MeAE values was obtained in the case of K-NN algorithm when compared to other algorithms, which were found to be 1127.238, 33.574, 22.766 and 12.194 respectively. However, MSE, RMSE, MAE and MeAE values were obtained in the case of linear regression were found to be high with values 2001.206, 44.735, 34.766 and 22.313 respectively.

Table 4 : Comparison of different machine learning algorithms for prediction of groundnut yield in summer season.

Models	MSE	RMSE	MAE	MeAE
Linear regression	73.488	8.573	6.790	6.586
LASSO	**60.632**	**7.787**	**6.916**	**6.476**
Ridge Regression	61.304	7.830	6.984	6.585
ELNET	61.517	7.843	7.055	6.880
SVR	63.065	7.941	6.557	6.501
KNN	60.838	7.801	6.983	6.512

Least values of MSE, RMSE, MAE and MeAE values was obtained in the case of LASSO algorithm when compared to other algorithms, which were found to be 60.632, 7.787, 6.984 and 6.776 respectively. However, high values of MSE, RMSE, MAE and MeAE were obtained in the case of linear regression with values 73.488, 8.573, 6.79 and 6.586 respectively. Also, except linear regression all other algorithms were found to be performing well for the dataset.

In the case of kharif groundnut yield prediction, K-NN method turned out to be best model with least values of MSE, RMSE, MAE and MeAE may be possibly because of its ability to handle the complex datasets and able to capture the minute patterns in the datasets. Similar results were obtained were by the study conducted by Shah and Shah (2018) in which groundnut crop yield was predicted using machine learning techniques. Four algorithms were evaluated namely, Multiple Linear Regression, Regression Tree, K-NN and Artificial Neural Networks out of which K-NN turned out to be best algorithm.

Similarly, in the case of summer groundnut yield prediction, LASSO algorithm turned out to be best algorithm based on weather parameters. Possibly, due to the application of shrinking coefficient which eliminates the variables which are least contributing on the dependent variable i.e., yield of grapes. Similar results were obtained by Kumar *et al.* (2019*),* in which LASSO and stepwise regression was compared for wheat yield prediction based on weather parameters.

Conclusion

The research revealed significant correlations between weather parameters and groundnut yield, offering crucial information for agricultural stakeholders. Specifically, the study highlighted the significantly positive impact of relative humidity and rainfall on kharif season yields, along with the significantly negative influence of temperature in kharif season whereas in summer season significantly negative impact of temperature on summer yield. Moreover, it identified LASSO regression as a promising model for predicting summer groundnut yields, showcasing its potential in optimizing crop management. Similarly, for kharif groundnut yield prediction K-NN turned to be the best model. These findings have far-reaching implications for sustainable agriculture and climate change adaptation in the region. They empower farmers and policymakers with actionable insights to enhance crop productivity and address food security challenges in the face of climate variability. Furthermore, this study underscores the pivotal role of advanced analytical tools in harnessing data-driven solutions for resilient and efficient agricultural practices, ensuring a more secure future for the agricultural sector in Dharwad district and beyond.

References

Anonymous, (2021). Directorate of Economics and Statistics (DES), Government of Karnataka. des.karnataka.gov.in.

Chlingaryan, A., Sukkarieh, S. and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and electronics in agriculture*, 151: 61-69.

Cho, S., Kim, H., Oh, S., Kim, K. and Park, T. (2009). Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. *BMC Proceed*, 3(7): 1-6

Enas, G.G. and Choi, S.C. (1986). Choice of the smoothing parameter and efficiency of k-nearest neighbor classification. Statistical Methods of Discrimination and Classification, 2(3): 235-244.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55-67.

Jasna, V.K., Som, S., Burman, R.R., Padaria, R.N. and Sharma, J.P. (2014). Socio economic impact of climate resilient technologies. International *Journal of agricultural and Food Sciences Technology*, 5(3): 185-190.

Kalarani, M.K., Senthil, A., Punitha, S., Sowmyapriya, S., Umapathi, M. and Geethalakshmi, V. (2023). Abiotic Stress Responses in Groundnut (*Arachis hypogaea L*): *Mechanisms and Adaptations in Legumes. Physiology and Molecular Biology of Abiotic Stress Tolerance*, 5(8): 279-290.

Karthikeya, H.K., Sudarshan, K. and Shetty, D.S. (2020). Prediction of agricultural crops using KNN algorithm. *International Journal of Innovative Science and Research Technology*, 5(5): 1422-1424.

Kumar, K.N.R., Satpathi, A., Reddy, M.J.M., Setiya Pand Nain, A.S. (2023). Development of groundnut yield forecasting models in relation to weather parameters in Andhra Pradesh, India. *Journal of Agrometeorology*, *25*(3):440-447.

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.

Lipiec, J., Doussan, C., Nosalewicz, A. and Kondracka, K. (2013). Effect of drought and heat stresses on plant growth and yield *International Agrophysics*, 27(4): 12-16.

Mohapatra, S. and Chaudhary, N. (2022). Statistical Analysis and Evaluation of Feature Selection Techniques and implementing Machine Learning Algorithms to predict the Crop Yield using Accuracy Metrics, *Engineered Science*, 21: 787.

Nageshwar, R., George, V. and Shivashankar, R. (2008). PFWD, CBR and DCP evaluation of lateritic subgrades of Dakshina Kannada, India

Palanivel, K. and Suriyanarayanan, C. (2019). An approach for prediction of crop yield using machine learning and big data techniques. *International Journal of Computer Engineering and Technology*, 10(3): 110-118.

Pavan, T. (2022). Statistical evaluation of the impact of weather parameters on selected rainfed crops of Chitradurga district, *M.Sc. (Agri) Thesis*, University of Agricultural Sciences, Dharwad, Karnataka, India.

Raza, A., Razzaq, A., Mehmood, S.S., Zou, X., Zhang, X., Lv, Y. and Xu, J. (2019). Impact of climate change on crops adaptation and strategies to tackle its outcome. *Plants*, 8(2): 34.

Reyer, C.P., Otto, I.M., Adams, S., Albrecht, T., Baarsch, F., Cartsburg, M., Coumou, D., Eden, Ludi E., Marcus, R. and Mengel, M. (2017). Climate change impacts in Central Asia and their implications for development. *Regional Environmental Change*, 17: 1639-1650.

Saravanakumar, V., Lohano, H.D. and Balasubramanian, R. (2022). A district-level analysis for measuring the effects of climate change on production of rice: evidence from Southern India *Theoretical and applied climatology,* 150(A4): 941- 953.

Shah, V. and Shah, P. (2018). Groundnut crop yield prediction using machine learning techniques *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3(5): 1093-1097.

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 Biometeorol*, 67(1): 165-180.

Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net *Journal of the Royal Statistical Society*, 67 (2): 301–302.