Original Research Article

Studying on crop response model for grapes under climate change scenario: statistical study approach.

Abstract:

Grapes vine originally a temperate fruit crop and it's also grown successfully under tropical conditions. Grape is one of the economically important fruit crops grown in India. As Theni district is the leading producer of grape in Tamil Nadu, followed by Coimbatore and Dindigul, this study is centred on the Theni-Kambum block area. In this region, Muskat Humbug is a well-liked cultivar that yields more than other varieties. In this study, this cultivar was employed. In this study, an artificial neural network (ANN), multiple linear regression (MLR), and elastic net (ELNET) regression methods were used to construct a yield prediction model (ANN). Additionally, we evaluate our model over a two-year period using field-level data from GRS and neighbouring farms. Finding the best-fit model for predicting grape yield and PDI using meteorological parameters in the Theni district is the goal of this communication. The model is chosen according to many performance indicators including RMSE, MAPE, MAE, and R^2. Among the three techniques developed in the study, the Artificial Neural Network is found to be best for prediction of grape yield based on weather and disease incidence for the available data in the studied region.

1.INTRODUCTION:

Grapes are produced on every continent, in temperate, subtropical, and tropical climates, and in a variety of agro-ecological settings ranging from mountains to plains to sea beaches. Grapes are produced in a range of temperatures and soils in India, with more than 80% of the land area lying within the tropical climatic region. India holds the distinction of having the greatest grape yield in the world. Grape suffers significant crop losses due to downy mildew, powdery mildew, and anthracnose. [1]. Plant disease prediction has evolved as a well-established component of epidemiology that is quickly being integrated into disease management. The mathematics of disease progression has advanced to the point of being a powerful and acknowledged component in epidemic management and prediction. [2]. S. Sannakki et al., 2013 forecast the weather using a modified k-NN technique and a Feed Forward Neural Network, and then use characteristics such as humidity and temperature to forecast disease outbreaks in grapes. Plant disease models have traditionally employed Leaf Wetness Duration (LWD) and temperature to forecast infection and colonisation, and subsequently determine the risks of an epidemic. These models have been used with observed climate data to monitor advantageous times, indicating control methods or strategies. [4]. Precipitation (availability of water) and air temperature are the primary meteorological elements that influence grape and wine quality and utilised an ensemble of CMIP6 model data to assess all possible changes in water availability in the area around Sevastopol by the middle and end of the twenty-first century for two shared Socio-economic Pathway scenarios (SSP2-4.5 and SSP5-8.5). [6]. A new technique for evaluating regional climate scenarios based on the statistical region model STAR has been developed. The approach improves applicability and reliability in viticultural elements and focuses on evaluating adaption measurements rather than predictions [3]. Multiple regression approach was used to develop an agro-climate grape yield (ACGY) model using climatic parameters and the developed model had been statistically tested for its predictive ability. Sensitivity analysis was carried out for the developed ACGY model using the parametric sensitivity method [7]. This study intends to develop a prediction model using yield as a dependent variable with climatic variables and percent disease incidence as independent variable and PDI as dependent variable with climatic variables as dependent variable.

2. MATERIALS AND METHODOLOGY:

2.1Study of data collection:

The daily weather parameters and yield were collected in Grape research station, Theni district for the years (2010-2021). The daily weather data are taken on an average to form year wise weather data. The primary data for validating the model had been collected from Theni and surrounding villages to farmers for two years (2021 and 2022).

2.2. Methodology:

2.2.1Multiple Linear Regression:

Multiple regression, or MLR, is a statistical method for predicting the outcome of a response variable by integrating a number of explanatory factors. The linear relationship between explanatory (independent) and response (dependent) variables is attempted to be represented using multiple linear regression. Multiple regression is simply an extension of ordinary least-squares (OLS) regression since it includes more than one explanatory variable.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + e$$

where, for *i*=*n* observations:

Y_i=dependent variable

X_i=explanatory variables

β_0=y-intercept (constant term)

β_p=slope coefficients for each explanatory variable

e =the model's error term (also known as the residuals)

2.2.2 Elastic Net Regression:

ELNET regression stands for elastic net regression, which is a mixture of penalties from LASSO and ridge regression (Hoerl and Kennard, 1970) that improves statistical model regularisation. During the regularisation procedure, the L1 component of the penalty generates a sparse model. The penalty's quadratic component (L2), on the other hand, makes the L1 portion more stable on the path to regularisation, removes the quantity limit of variables to be picked, and promotes the grouping effect. As a result, it reduces the impact of certain aspects but not completely eradicating them (Cho et al., 2009).

$$\hat{\beta}(Enet) = \left(1 + \frac{\lambda_2}{n}\right)\left\{arg\ min_{\beta}\|y - X\beta\|^2 + \lambda_1\|\beta\| + \lambda_2\|\beta\|^2\right\}$$

where, λ_1 and λ_2 are LASSO and ridge regression penalities.

The lambda values with the lowest average mean squared error were chosen using cross-validation with leave-one-out (Piaskowski et al., 2016). The overall strength of the penalty is controlled by tuning parameter λ (Hastie and Qian, 2014). Analysis of the data was performed using the R package 'glmnet' (Friedman et al., 2009).

2.2.3 Artificial Neural Network:

A neural network is a massively parallel network of linked basic processors (neurons), each of which accepts a set of inputs from other neurons and computes an output, which is transmitted to the output nodes. A neural network may therefore be represented in terms of individual neurons, network connectivity, weights associated with neuron interconnections, and neuron activation function. The neuron gets a set of n inputs from its neighbours, x_i, I = 1,2,...,n, as well as a bias of one. Each input is connected with a weight (w_i).The weighted sum of the inputs determines the state or activity of a neuron and is given by

$$a = \sum_{i=1}^{n+1} w_i x_i = W^T X$$

Where, $X = \{x_1 x_2 \dots x_n\ 1\}^T$. The output of the neuron is commonly described by a sigmoid function as

$$f(a) = \frac{1}{1 + e^{-a}}$$

2.3Model performance metrices :

The performance of the statistical models was evaluated using coefficient of determination (R^2), Root mean squared error (RMSE), mean absolute percentage error (MAPE), and Mean absolute error (MAE) by following formulas:

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(Y_j - \hat{Y}_j)^2}{\sum_{j=1}^{n}(Y_j - \bar{Y}_j)^2}$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(Y_j - \hat{Y}_j)^2}{n}}$$

$$MAPE = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{Y_j - \hat{Y}_j}{Y_j}\right| * 100$$

$$MAE = \frac{\sum_{j=1}^{n}|Y_j - \hat{Y}_j|}{n}$$

where, Y_j – Actual yield, $\hat{Y}_j$ – Model yield respectively, n-number of years.

2.4 Correlation:

Correlation coefficients are used to quantify the strength of a linear association between two variables, x and y. A linear correlation coefficient greater than zero indicates a positive relationship. A number less than zero indicate a negative relationship. Finally, a value of 0 denotes that the variables x and y are unrelated.

$$\text{Correlation} = \rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

In this study, the dependent variable is nut yield and the independent variables are plant height, stem girth, Female flowers in inflorescence, husk Thickness, copra content, minimum Temperature, rainfall, and relative humidity. Correlation is carried out between nut yield and all other factors to find the factors which are all highly responsible for multi collinearity between predictor variables.

3. ANALYTICAL RESULT FOR THREE DIFFERENT TECHNIQUES:

The intercept and coefficients of multiple linear regression models are shown in the Table 1. The fitted models in the table revealed that the variables relative humidity and minimum temperature had positive impact on grape yield and the variables such as, rainfall, PDI and maximum Temperature were negatively impact on grapes yield .The Actual and predicted yield by multiple linear regression are shown in the Table 2. The Intercept and coefficients of PDI models are shown in the Table 3.The fitted models in the table revealed that the variables minimum temperature and relative humidity showing positive impact on PDI and the other variables like maximum temperature and rainfall showing slight negative impact on

PDI. The Actual and predicted yield by PDI are shown in the Table 4.The Intercept and coefficients of elastic net regression models are shown in the Table 5.Observed and Predicted value of ELNET model are shown in the Table 6. Intercept and coefficients of elastic net regression with PDI are shown in the Table 7. Observed and predicted value of ELNET model with PDI are shown in the Table 8. Predicted yield by ANN are shown in the Table 9.Observed and predicted value of ANN model for the years (2010-2021) are shown in the Table10.

Table 1. Intercept and coefficients of multiple linear regression models:

Independent Variables	Reg. coefficients (b)	Standard Error (SE(b))	T Test	p-value
Intercept	41.75484	29.00337	1.439655	0.200027
Maximum Temperature	-2.314**	0.877	-2.637	0.037755
Minimum Temperature	2.843*	1.235	2.303	0.059968
Relative Humidity	0.071	0.112	0.632	0.536572
Rainfall	-0.044**	0.015	-2.958	0.024579
PDI	-0.786	0.506	-1.552	0.169898

Table 2. Actual and predicted yield by multiple linear regression:

Year	Actual Yield	Predicted Yield
2010	25.27	23.78114
2011	24.38	23.16853
2012	24.79	24.93047
2013	25.00	23.95473
2014	23.53	24.39819
2015	22.19	23.4507
2016	26.75	26.50789
2017	24.49	24.14511
2018	22.15	22.63229
2019	22.58	24.3132
2020	27.44	27.41715
2021	20.12	20.26463

Table 3. Intercept and coefficients of PDI models:

Independent Variables	Reg. coefficients (b)	Standard Error (SE(b))	T Test	p-value
Intercept	10.40915	21.34918	0.487567	0.640755
Maximum Temperature	-0.216	0.650	-0.332	0.745509
Minimum Temperature	0.306	0.914	0.335	0.750315
Relative Humidity	0.013	0.084	0.153	0.877668
Rainfall	-0.002	0.011	-0.157	0.875496

Table 4 . Actual and predicted yield by PDI:

Year	Actual Yield	Predicted Yield
2010	10.32	10.83898
2011	11.38	10.8317
2012	10.41	10.86152
2013	10.47	10.95588
2014	10.86	10.85827
2015	9.58	10.76451
2016	12.14	11.13124
2017	10.63	10.95449
2018	10.79	10.80761
2019	11.52	11.03571
2020	10.74	11.24892
2021	12.45	10.7523

Table 5. Intercept and coefficients of elastic net regression models:

(Intercept)	7.595969
Maximum Temperature	0
Minimum Temperature	0
Relative humidity	0.114039
Rainfall	-0.0018
PDI	0.84348

Table 6.Observed and Predicted Value of ELNET model:

Year	Observed Value	Predicted Value
2010	25.27	24.20834
2011	24.38	24.31131
2012	24.79	24.11563
2013	25.00	24.92588
2014	23.53	24.44693
2015	22.19	23.32742
2016	26.75	26.57449
2017	24.49	25.05627
2018	22.15	25.32153
2019	22.58	25.37917
2020	27.44	24.96572
2021	20.12	25.92124

Table 7. Intercept and coefficients of elastic net regression with PDI:

(Intercept)	0.18477306
Maximum Temperature	-2.27533478

Minimum Temperature	4.12084136
Relative humidity	-0.10271152
RF	-0.01797528

Table 8. Observed and predicted value of ELNET model with PDI:

Year	Observed Value	Predicted Value
2010	10.32	10.21824
2011	11.38	11.35727
2012	10.41	10.44909
2013	10.47	10.52533
2014	10.86	10.9253
2015	9.58	9.579368
2016	12.14	12.10541
2017	10.63	10.40887
2018	10.79	7.799166
2019	11.52	12.44111
2020	10.74	13.77059
2021	12.45	8.911966

3.1 Artificial Neural Network:

A mathematical model that attempts to imitate the structure and capabilities of biological neural networks is known as an Artificial Neural Network (ANN). Every artificial neural network starts with an artificial neuron, which is a simple mathematical model (function). A model contains three basic sets of rules: multiplication, summation, and activation. The inputs are weighted at the entry of the artificial neuron, which implies that each input value is multiplied by an individual weight. The sum function in the centre region of the artificial neuron adds all weighted inputs and bias. At the exit of an artificial neuron, the total of previously weighted inputs and bias passes through an activation function, also known as a transfer function.

The processing element is split into two parts. The weighted inputs are simply aggregated in the first portion; the transfer function, sometimes referred to as the activation function, or second part, is essentially a non - linear filter. The output values of an artificial neuron are constrained or compressed by the activation function to a region between two asymptotes. The sigmoidal function is the most often used function.

Table 9. Predicted yield by ANN:

Parameter Estimates			
Predictor		Predicted	
		Hidden Layer 1	Output Layer
		H(1:1)	Yield
Input Layer	(Bias)	-0.117	
	MAXT	-0.921	
	MINT	0.370	
	RH	1.581	
	RF	-2.112	
	PDI	0.497	
Hidden Layer 1	(Bias)		0.351
	H(1:1)		0.706

Table10. Observed and predicted value of ANN model for the years (2010-2021).

YEAR	Observed value	Predicted Value
2010	10.32	10.41
2011	11.38	11.38
2012	10.41	10.33
2013	10.47	10.52
2014	10.86	11.20
2015	9.58	10.74
2016	12.14	12.14
2017	10.63	10.43
2018	10.79	10.44

2019	11.52	11.50
2020	10.74	10.99
2021	12.45	12.45

3.2 COMPARITIVE RESULT:

The yield and PDI developed models were compared using different performance metrics. Based on the comparison, ANN model for both yield and PDI developed models was performing far better than other models with high R-squared and low RMSE, MAE and MAPE values which are followed by ELNET and MLR.

Table 11.Comparison of Yield Model:

	R-squared	RMSE	MAPE	MAE
MLR	0.88	0.821824	2.41272	0.545572
ELNET	0.819305	3.405615	13.37552	2.962498
ANN	**0.983076**	**1.032531**	**4.378608**	**0.988**

Table 12.Comparison of PDI model:

	R-squared	RMSE	MAPE	MAE
MLR	0.574	0.834425	5.15888	0.606602
ELNET	0.879	1.261784	18.88606	2.14034
ANN	**0.987**	**0.009016**	**1.525314**	**0.164**

4. RESULT AND DISCUSSION:

Figure 1 explains that the variables were selected based on the correlation coefficient between yield and other variables. Like that, we had selected Minimum Temperature, Relative humidity and Rainfall as weather parameters for the same aspects. From figure 2, yield as a function of years, it is evident that the yield was peak at 2019-2020 thereafter it was decreased and reached lowest yield (20.12 t/ha) in 2020-2021. During 2020-2021 the disease incidence was increasing trend and having a value of (12.45%). In figure 3, the percentage disease incidence is increases as increasing the relative humidity and rain fall.

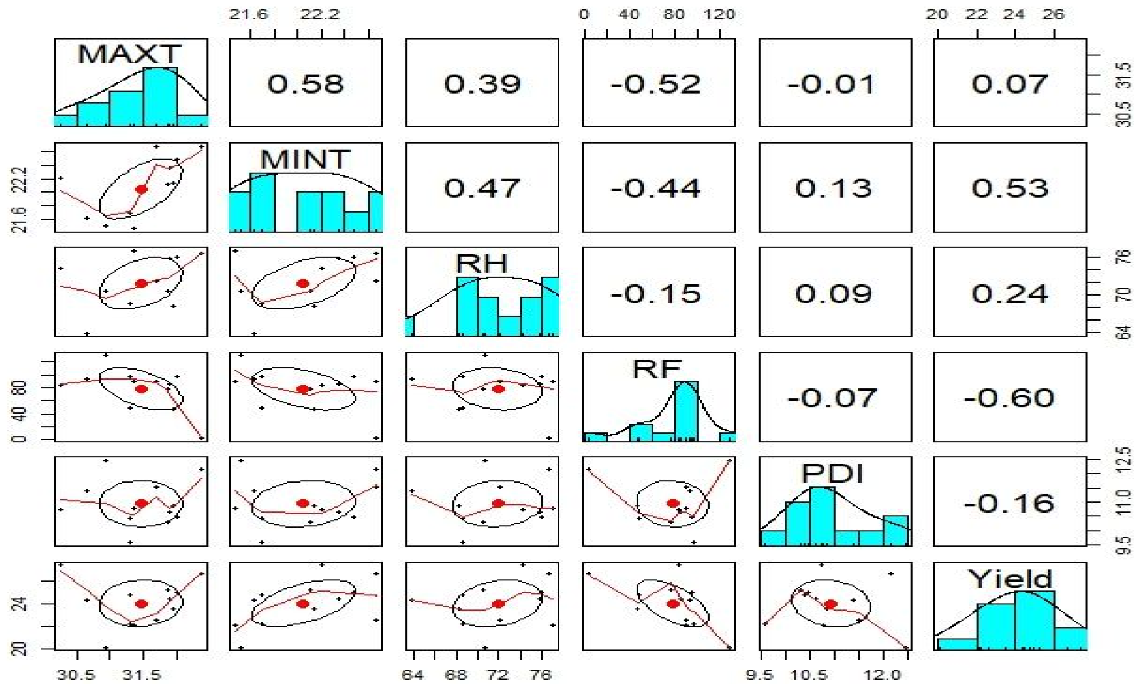

Fig 1. Correlogram showing correlation between yield and climatic factors.

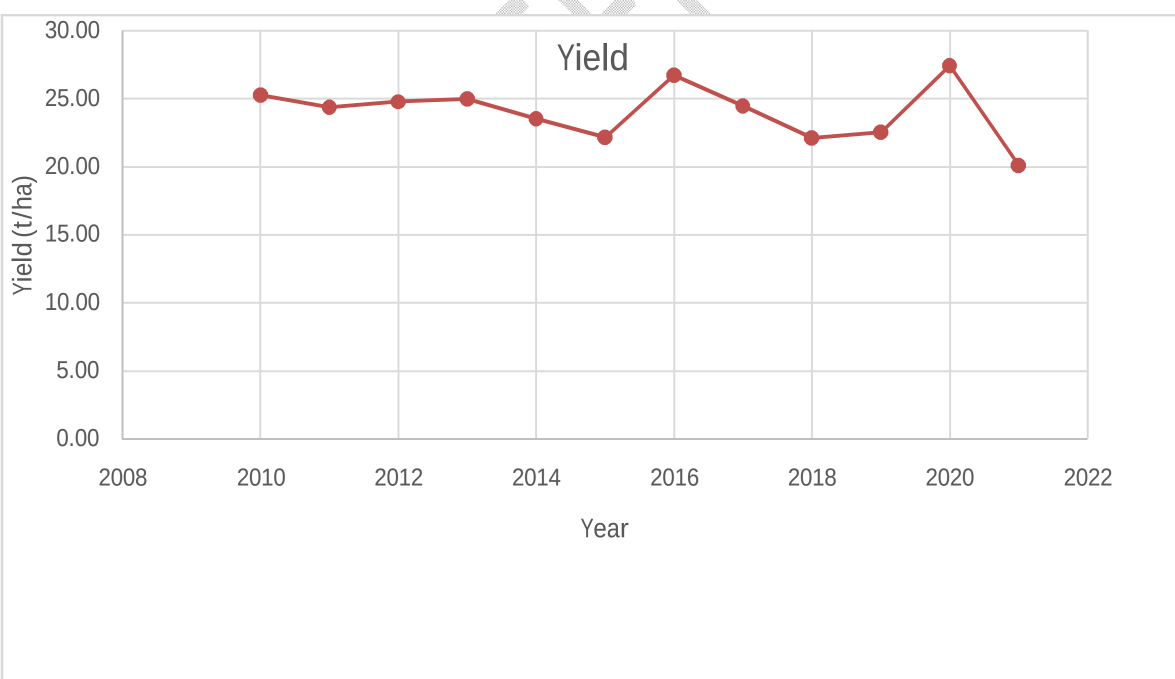

Fig 2: Comparison of observed and predicted yield as a function of years for various values of climatic factors. The line represent the predicted yield and dot repreent the observed yield for Theni and surrounding villages.

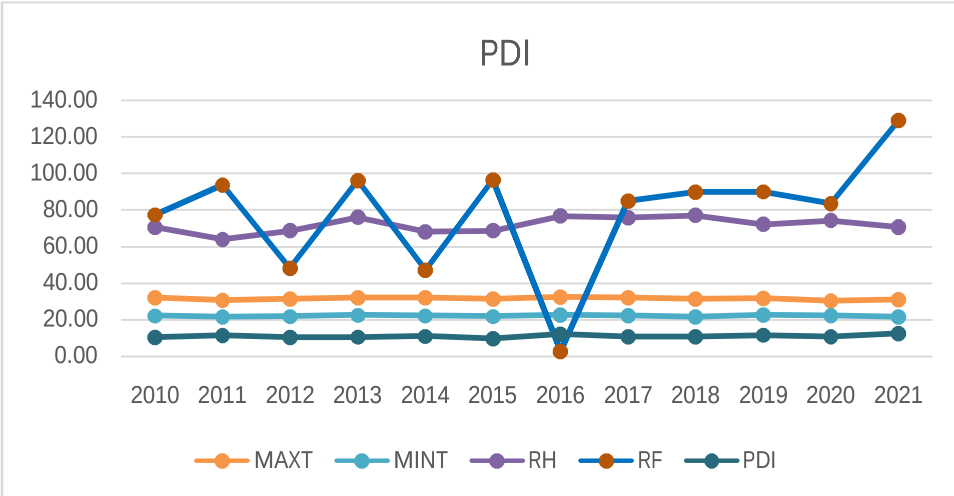

Fig 3: Comparison of observed and predicted PDI as a function of years for various values of climatic factors. The line represent the predicted PDI and dot represent the observed PDI for Theni and surrounding villages.

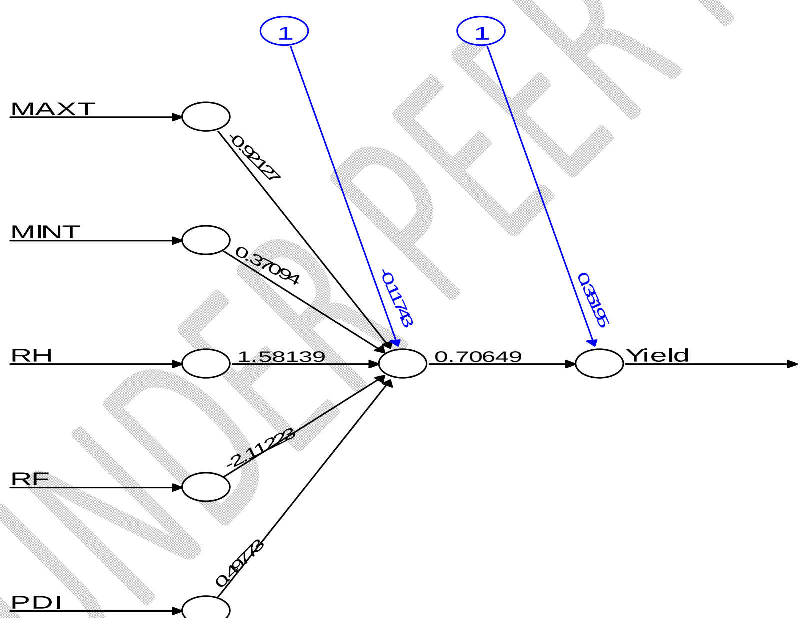

Fig 4. Graphical representation of ANN.

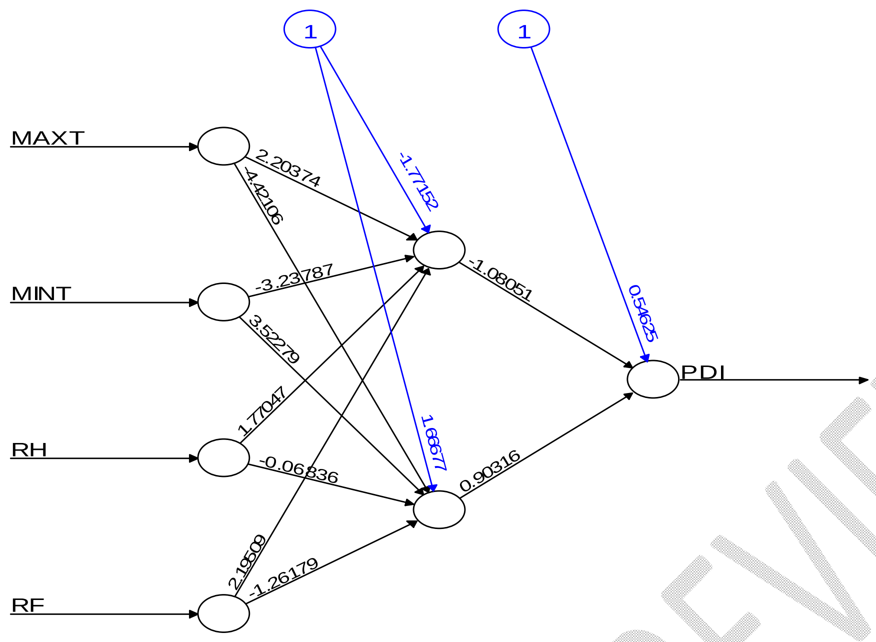

Fig 5. Graphical representation of ANN model with PDI

The number of principal components retained for this study is 4 which explain 90 percent of the variation in the data. The retained components are then used as an input layer to train Artificial Neural Network. The number of hidden layers selected is 1 and the optimum number of neurons in the hidden layer is 2. The network flow is represented in Figs. 4 and 5 and the connection weights are displayed.

5. CONCLUSION:

From the study, three different statistical techniques were used and the techniques were compared by using different error measures. It is concluded that the ANN was found to be the best technique to predict the grape yield. This analytical result helps for the better understanding of the yield prediction for grape. Through numerical experiments we have been able to get further insight into thresholds for disease extinction that can contribute to crucial knowledge of disease control.

REFERENCES:

1. A report of the expert consultation on viticulture in Asia and the Pacific. May 2000, Bangkok, Thailand. RAP publication:2000/13.
2. Seem, R., 2001. Plant disease forecasting in the era of information technology. In: Plant Disease Forecast: Information Technology in Plant Pathology, Kyongju, Republic of Korea.
3. Stock, M., Gerstengarbe, F.W., Kartschall, T. and Werner, P.C. (2005). Reliability of climate change impact assessments for viticulture. Acta Hortic. 689, 29-40

4. Jabrzemski, R., Sutherland, A., 2006. An innovative approach to weather-based decision-support for agricultural models. In: 22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology. American Meteorological Society, Washington, DC, USA.
5. S. Sannakki, V. S. Rajpurohit, F. Sumira and H. Venkatesh, 2013. A neural network approach for disease forecasting in grapes using weather parameters. Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1-5.
6. Vyshkvarkova E, Rybalko E, Marchukova O, Baranova N. Assessment of the Current and Projected Conditions of Water Availability in the Sevastopol Region for Grape Growing. *Agronomy*. 2021; 11(8):1665.
7. S.J. KADBHANE, & V.L.MANEKAR. (2020). Development of agro-climatic grape yield model for Nashik region, Maharashtra, India. *Journal of Agrometeorology*, *22*(4), 494–500.
8. Cho, S., Kim., Oh, S., Kim, K., and Park, T. 2009. Elastic net regularization approaches for genome wide association studies of rheumatoid arthiritis. *BMC proceedings.* 3(7):1-6.
9. Hoerl, A. E., and Kennard, R. W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55-67.
10. Friedman, J., Hastie, T., and Tsbshirani, R. 2009. Glmnet: Lasso and elastic net regularized generalized linear models. *R package version.,* 1(4): 1-24.
11. Piaskowski, J. L., Brown, D., and Campbell, K.G. 2016. Near infrared calibration of soluble stem carbohydrates for predicting drought tolerance in spring wheat. *Agronomy Journal.,* 108(1): 258-293.

Appendix:

ANN programme using R Code:

```
> set.seed(132)
> data<-data.frame(data)
> train_id<-data[1:7,]
> test_id<-data[8:9,]
> max = apply(data , 2 , max)
> min = apply(data, 2 , min)
> scaled = as.data.frame(scale(data, center = min, scale = max - min))
> trainNN = scaled[1:7,]
> testNN = scaled[8:12,]
> library(neuralnet)
> n<-neuralnet(Yield~.,data = trainNN,hidden=1,linear.output = TRUE)
> n$result.matrix
>plot(n)
> predict_testNN1 = compute(n, testNN[,-1])
```

```
> predict_testNN1
> predict_testNN = (predict_testNN1$net.result * (max(data$Yield) - min(data$Yield))) + min(data$Yield)
> predict_testNN
> library(Metrics)
> mae(actual,predicted)
> mape(actual,predicted)
> rmse(actual, predicted)
```