https://doi.org/10.56093/JISAS.V79I2.9
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 79(2) 2025 141–147

Artificial Neural Networks Models for the Classification of Genotypes for Rice Yield

Shavi Gupta, Manish Sharma, Sushil Gupta, Nishant Jasrotia and Bupesh Kumar
Sher-e-Kashmir University of Agricultural Sciences & Technology-Jammu, Jammu

Received 17 September 2025; Revised 27 November 2025; Accepted 01 December 2025

SUMMARY

Artificial neural networks, or ANNs, are widely used in many scientific fields like classification, image processing, identification, and prediction. This research was done at the SKUAST Jammu for classification of 140 rice varieties on the basis of their yield. Classification was made in terms of 9 morphological characters such as yield per plant, number of days for 50percent flowering, number of days for full flowering, plant height, number of effective tillers per plant, panicle length, grain length, grain width and ratio of grain length & grain width. The yield per plant was grouped into three categories and used as the response variable, while the remaining traits served as predictor variables. The rice genotypes were then classified according to neural networks (NN) methods Multilayer Perceptron (MLP) and Radial Basis Function (RBF). The ability measures of classification such as Accuracy Rate, Kappa Statistics, Average Precision, Average Recall and F1 Score were used for testing samples**.** It is observed that RBF NN performed better than MLP NN for different classes of yield on the basis of classification ability measures. The variable effective number of tillers per plant found out to be important variable as per RBF NN whereas plant height was important variable in case of MLP NN.

Keywords: Classification; Rice; Artificial Neural Network; Multilayer Perceptron; Radial basis Function; Classification ability measures.

1. INTRODUCTION

Everywhere in the world, but particularly in East and South Asia, Latin America, India, and Iran, rice (Oryza sativa L.) is a staple food. The Poaceae family includes rice as a grass cereal. This study looked at the quality of products based on their outward structure and morphological characteristics in order to define rice (Jayas *et al.,* 2000; Harper *et al.,* 1970). Classification is one of the important areas of research in the field of data mining and neural network is one of the widely used techniques for classification. One data mining technique for predicting an object's class is classification. This depicts supervised learning. categorization predicts a label (ordered, distinct) and classifying data requires two processes. (i) A classifier is constructed to characterize a predefined set of data classes in the first stage, known as the learning step (training step). (ii) The test data are used to estimate the classifier accuracy, and the model constructed in the previous step is utilized to classify unknown data. Numerous techniques exist for classification, such as artificial neural networks, decision trees, K nearest neighbours, and naïve Bayesian classifiers. Chen *et al.,* (2010) used neural networks and pattern recognition algorithms to identify five corn types with an accuracy of above 90 percent. Galdon *et al.,* (2010) showed that eight onion cultivars of varying colors (white, golden, red) were studied under uniform growing conditions. Cluster and principal component analyses revealed that bulb color, not variety, had the strongest influence on flavonoid, phenol, and pungency profiles. Artificial neural networks were employed to assess the feasibility of variety discrimination. Pazoki & Pazoki (2011) described a method for categorizing five cultivars of rain-fed wheat grain using an artificial neural network. According to the experiment results, the average accuracy was 86.48 percent before the UTA

Corresponding author: Manish Sharma
E-mail address: manshstat@gmail.com

algorithm's feature selection application, and after that it climbed to 87.22 percent. Sharma *et al.,* (2022) used Discriminant function analysis to develop price models for tomato crop in Jammu region. The study revealed that price models for both summer and winter seasons are significant. Gupta *et al.,* (2023) use statistical techniques like discriminant analysis and artificial neural networks (ANN), such as multilayer perceptron neural networks for various yield classes, to categorize the rice genotypes. These methods were used on primary data acquired from the SKUAST, Jammu trial for 100 genotypes of rice for five morphological factors. For the categorization of genotypes for various classes of rice genotype yield, the Artificial Neural Network model (85%) outperformed Discriminant analysis (75%). For the primary data pertaining to 140 rice genotypes from the SKUAST, Jammu trial based on maturity, Gupta *et al.* (2025) employed statistical and artificial neural network models. On the basis of different evaluation measures neural network models outperformed statistical models. Our study was designed to compare Artificial Neural Network (ANN) models with traditional statistical techniques, specifically Ordinal Logistic Regression and Discriminant Analysis. Consequently, methods such as Support Vector Machines (SVM) and Random Forest were excluded from consideration. Furthermore, in the walnut study by Jasrotia *et al.,* (2025), both linear and radial SVM approaches demonstrated limited efficiency, whereas ANN methods consistently outperformed them. On the basis of different evaluation measures neural network models outperformed statistical models.

In the present paper, two artificial neural networks have been applied for classification of rice genotypes on the basis of their yield and their performances are compared.

2. MATERIAL AND METHODS

The primary data for various rice genotype characteristics, including yield per plant (X1:classes), number of days for full flowering (X2), plant height (X3), number of effective tillers per plant (X4), panicle length (X5), grain length (X6), grain width (X7), ratio of grain length & grain width (X8), and number of days for 50% flowering (X9) of 140 rice genotypes from the trail laid in SKUAST Jammu is used. The yield per plant of rice genotypes is considered as dependent variable which has been further classified into three categories by using Jenks natural breaks optimization method (Jenks, 1967) as

Variable (classes)	Classes of Genotypes	Number of Days
Yield per plant(X1)	Low Yield(1)	<36 grams
	Medium Yield (2)	36 grams – 59 grams
	High Yield (3)	>59 grams

and rest of the characters related to rice plants are considered as explanatory variables for the study. The data set is randomly divided into training and testing datasets. Classification was made in terms of the eight characters and fed to the multilayer perceptron (MLP) and radial basis function (RBF). An MLP (Multilayer Perceptron) network is a widely used network with a variety of uses. An MLP network's architecture consists of an input layer, one or more hidden layers, and one output layer. A training algorithm, like back propagation, is used to train a network once its structure has been created. By adjusting the weights and bias, the training method reduces error. Another important kind of neural networks are radial basis functions (RBF), where the activation unit is determined by the separation between the input and prototype vectors. Techniques for precisely interpolating a set of data points in a multidimensional space are the foundation of radial basis functions. All the input vectors had to be precisely mapped onto the matching target vectors in order to solve the exact interpolation problem. Due to the ability of a hidden unit to disperse its effect over the whole input space, MLPs are referred to as distributed-processing networks. On the other hand, because a hidden unit's effect is typically focused in a narrow area centred at the weight vector, Gaussian RBF networks are referred to as local-processing networks.

The statistical measures like accuracy rate, kappa statistics, average precision, average recall and F1 score to evaluate the classification performance the above mentioned artificial neural network models. These were calculated with the help of confusion matrix

Classes	A	B	c
a	Aa	Ab	ac
b	Ba	Bb	bc
c	Ca	Cb	cc

where a, b and c: are three classes; aa, bb & cc are the number of samples that are correctly classified as class a, b & c respectively; ab: number of samples that are in class a but predicted in class b; and the rest have the same explanation. The above-mentioned measures are given as $\text{Accuracy Rate} = \frac{\text{correctly classified data}}{\text{total data}} \times 100$

and $\text{Kappa Statistics} = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} x_{ir} x_{ic}}{N^2 - \sum_{i=1}^{k} x_{ir} x_{ic}}$, here, x_{ii} represents the sum of the diagonal elements of the confusion matrix, x_{ir} and x_{ic} denote the row and column totals of the matrix, respectively, and N is the total number of observations.

$$Avg.\,precision = \left(\frac{aa}{aa+ab+ac}\right) + \left(\frac{bb}{ba+bb+bc}\right) + \left(\frac{cc}{ca+cb+cc}\right)$$

and

$$Avg.recall = \left(\frac{aa}{aa+ba+ca}\right) + \left(\frac{bb}{ab+bb+cb}\right) + \left(\frac{cc}{ac+bc+cc}\right).$$

$$\text{Fl Score} = 2\frac{Avg.Precision * Avg.recall}{Avg.\,precision + AVg.recall}$$

3. RESULTS AND DISCUSSION

The descriptive summary of all variables used for classifying rice genotypes, including their mean values, standard deviation, coefficient of variation, minimum and maximum values, is presented in Table 1. The yield of the genotypes varies, with a minimum of 110 grams per plant and a maximum of 555 grams per plant. Therefore, classification is crucial to identify genotypes based on yield similarity. Additionally, the mean values of parameters X3, X6, and X8 show that their values are more than eight times greater than the standard deviations, indicating consistency among these variables for genotype classification. For other variables, two to three times greater than the standard deviations. The table also reveals that variable X1 has the highest coefficient of variation (27.88 percent), followed by X5, while variable X6 has the lowest coefficient of variation (12.56 percent), followed by X3. This indicates significant variation among the variables.

Table 1. Descriptive Statistics for all characters of Rice Genotypes

Variable (Unit)	Mean	Standard Deviation	CV (in percent)	Minimum	Maximum
X_1 (in grams)	49.120	13.697	27.884	110.000	555.000
X_2 (in days)	79.860	13.473	16.871	54.000	104.000
X_3 (in cm)	135.271	19.238	14.221	80.670	176.330
X_4 (in number)	23.690	5.763	24.327	11.000	38.000
X_5 (in cm)	28.667	7.825	27.296	15.750	50.000
X_6 (in mm)	9.025	1.134	12.565	6.384	11.594
X_7 (in mm)	1.786	0.385	21.556	1.006	2.981
X_8 (in ratio)	5.253	1.147	21.835	3.022	9.978
X_9 (in days)	77.170	13.804	17.887	50.000	106.000

The dataset has been split randomly into training and testing sets, with the main goal of the study being to assess ANN models for rice genotype classification. For training dataset 70 percent of the data is used and remaining 30 percent for testing dataset. Model performance was further validated using a 6-fold cross-validation approach to ensure robustness. Table 2&3 represents the information about the multilayered perceptron NN architecture and Radial Basis Function NN architecture respectively for yield of rice. The topological structure of MLP model consisted of 8 neurons in the input layer; 3 neurons (Low Yield, Medium Yield & High Yield classes) in the output layer and one hidden layer with 7 neurons as illustrated in Table 3. While in case of RBF model, 8 neurons in input layer, 10 neurons in hidden layer and 3 neurons in output layers depicted by Table 3.

The architecture of the MLP NN and RBF NN have been shown in the Fig. 1 and Fig. 2 respectively for yield of rice, light color lines display weights having more than zero, and the dark color lines show weight less than zero.

Table 2. Network Information Summary of MLP NN for Yield of Rice

Input Layer	**Covariates**		Number of days for full flowering(X2) Plant Height(X3) Number of effective tillers per plant(X4) Panicle Length(X5) Grain Length(X6) Grain Width(X7) Ratio of grain length and grain width(X8) Number of days for 50percent flowering(X9)
	Number of Units[a]		8
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	**Number of Hidden Layers**		1
	Number of Units in Hidden Layer 1[a]		7
	Activation Function		Hyperbolic tangent
Output Layer	**Dependent Variables**	1	Classes
	Number of Units		3
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

Table 3. Network Information Summary of RBF NN for Yield of Rice

Input Layer	**Covariates**		Number of days for full flowering(X2) Plant Height(X3) Number of effective tillers per plant(X4) Panicle Length(X5) Grain Length(X6) Grain Width(X7) Ratio of grain length and grain width(X8) Number of days for 50 percent flowering(X9)
	Number of Units		8
	Rescaling Method for Covariates		Standardized
Hidden Layer	**Number of Units**		10[a]
	Activation Function		Softmax
Output Layer	**Dependent Variables**	1	classes
	Number of Units		3
	Activation Function		Identity
	Error Function		Sum of Squares
a. Determined by the testing data criterion: The "best" number of hidden units is the one that yields the smallest error in the testing data.			

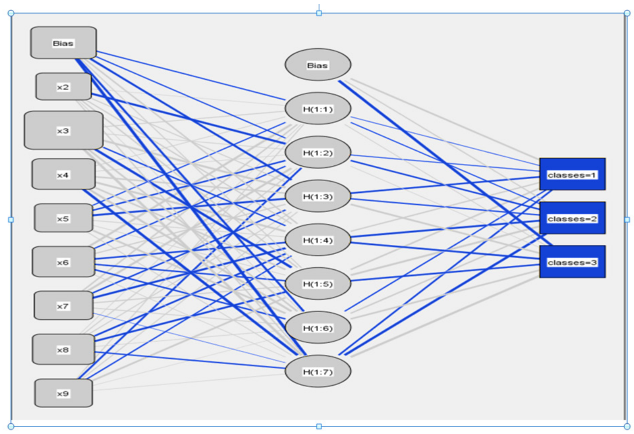

Fig. 1. Architecture of MLP NN for yield of Rice

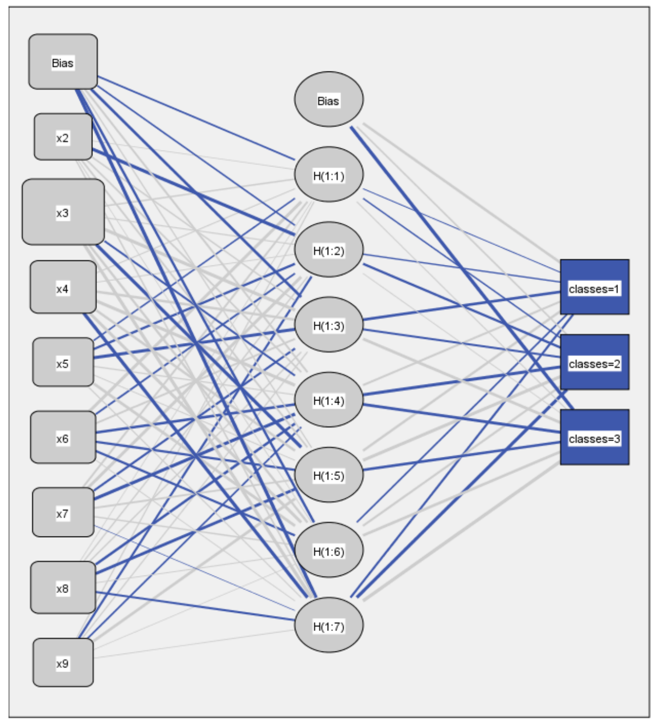

Fig. 2. Architecture of RBF NN for yield of Rice

Table 4 shows data on the MLP NN and RBF NN training and testing sample application of the final network. During training, the error function seeks to minimize error. For every category, the cross entropy error will have a predicted value, and each predicted value represents the likelihood that the example falls into one of the classes. When comparing the percentage of inaccurate predictions between the training and testing data sets (40.8 percent and 38.1 percent, respectively), it can be shown that the MLP NN misses matches with the original observed classes. While in case of RBF NN, the sum of squares error was calculated and percentage of incorrect predictions for training & testing datasets found out to be 32.7 percent & 23.8 percent respectively.

Table 4. Model Summary of MLP NN and RBF NN for Yield of Rice

Training	Cross Entropy Error	72.168
	Sum of Squares Error	**21.455**
	Percent Incorrect Predictions	40.8percent**(32.7percent)**
Testing	Cross Entropy Error	33.059
	Sum of Squares Error	**11.799**[a]
	Percent Incorrect Predictions	38.1percent**(23.8percent)**
Dependent Variable: classes		
a. The number of hidden units is determined by the testing data criterion: The "best" number of hidden units is the one that yields the smallest error in the testing data.		

Fig. 3 and Fig. 4 displays the importance of an independent variable according MLP NN and RBF NN. According to MLP NN, the variable plant height (100 percent), number of effective tillers per plant (40.5 percent), grain length (37.9 percent), and ratio of grain length & grain width (37.0 percent) were the most influential in classifying genotypes across different rice yield categories. While in case of RBF NN, the variable the number of effective tillers per plant (100 percent) and the plant height were key contributors.

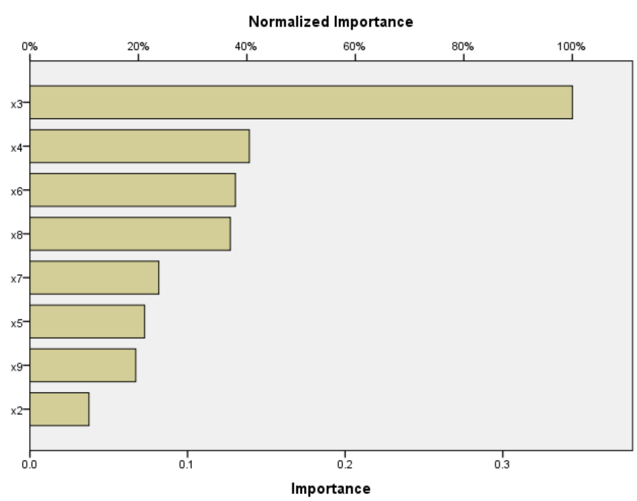

Fig. 3. Normalized independent variable importance(MLP NN)

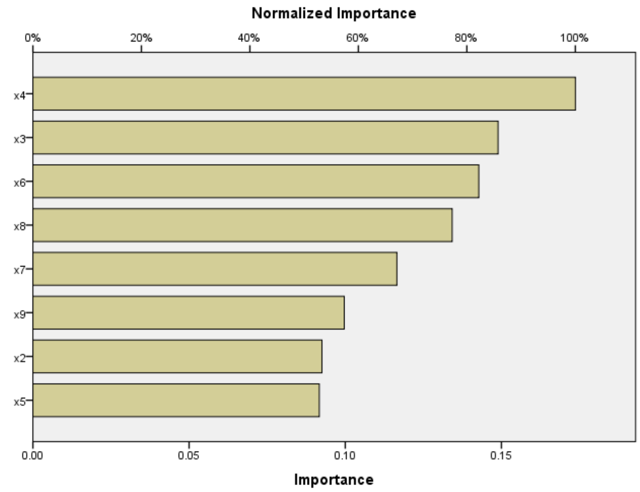

Fig. 4. Normalized independent variable importance (RBF NN)

In case of RBF NN, the number of effective tillers per plant (100 percent), plant height (85.8 percent), grain length (82.2 percent), ratio of grain length & grain width (77.3 percent), grain width (67.1 percent) and number of days for 50 percent flowering (57.4 percent) were have the greatest effect on classification of genotypes for different classes of yield of rice.

The ROC (Receiver Operating Characteristic) curve obtained for MLP NN and RBF NN were represented by Fig.5(a) and (b) respectively. According to MLP NN, which takes into account the information from combined training and testing samples, the low yield class has the highest area under the curve, followed by the medium and high yield classes. Whereas, on the basis of RBF NN, this shows that the low yield class has the largest area under the curve, followed by the medium and high yield classes.

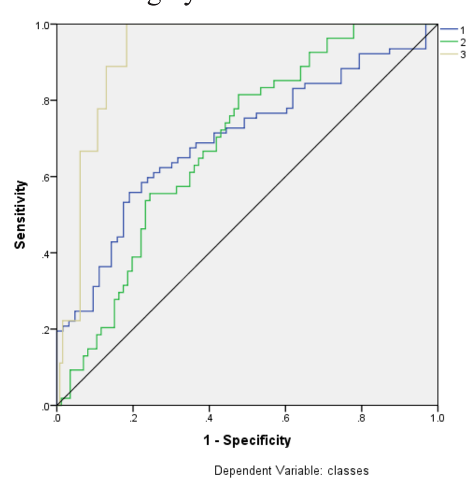

Fig. 5(a). ROC curve of MLP NN for yield of rice genotypes

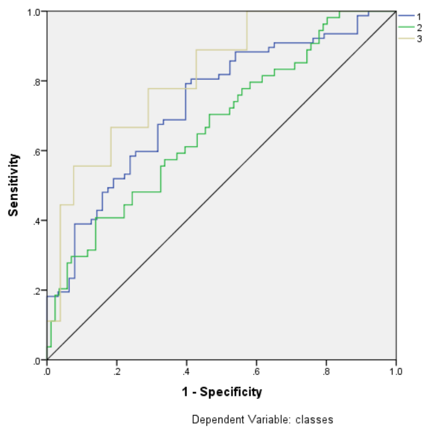

Fig. 5(b). ROC curve of RBF NN for the yield of rice genotypes

Table 5 presents the classification performance of the MLP NN model with respect to yield categories of rice genotypes. In the training dataset, the model correctly classified 43 of 56 low-yield genotypes, 14 of 37 medium-yield genotypes, and 1 of 5 high-yield genotypes, achieving an overall classification accuracy of 59.2 percent. In the testing dataset, correct classification occurred for 19 of 21 low-yield genotypes, 4 of 17 medium-yield genotypes, and 3 of 4 high-yield genotypes. The testing accuracy reached 75 percent, with 61.9 percent of the total testing cases classified correctly.

Table 5. Classification Matrix of MLP NN for Yield of Rice

Sample	Observed	Predicted			
		Low	Medium	High	Percent Correct
Training (70percent)	Low	43	10	3	76.7
	Medium	23	14	0	37.8
	High	4	0	1	25.0
	Overall Percent	71.4	24.4	4.1	59.2
Testing (30percent)	Low	19	1	1	90.4
	Medium	11	4	2	23.5
	High	0	1	3	75.0
	Overall Percent	71.4	14.2	14.2	61.9

The classification matrix of RBF NN for yield of rice genotypes is presented in table 6 which describes that in training dataset, 47 out of the 56 low yield genotypes, 17 out of the 37 Medium yield genotypes & 2 out of 5 high yield genotypes are correctly classified and in total, 67.3 percent of the training cases are classified correctly. While in testing data set, 18 out of 21 low yield mature genotypes, 13 out of 17 medium yield genotypes & 1 out of 4 high yield genotypes are correctly classified, and in total, 76.2 percent of the test cases were classified accurately.

Table 6. Classification Matrix of RBF NN for Yield of Rice

Sample	Observed	Predicted			
		Low	Medium	High	Percent Correct
Training (70percent)	Low	47	7	2	83.90
	Medium	19	17	1	45.90
	High	0	3	2	40.00
	Overall Percent	67.30	27.50	5.10	67.30
Testing (30percent)	Low	18	2	1	85.70
	Medium	4	13	0	76.40
	High	1	2	1	25.00
	Overall Percent	54.70	40.50	4.70	76.20

Table 7. Classification Ability of models for Yield of Rice for testing dataset

Criteria	Measures	Multilayered Perceptron Neural Network (MLP NN)	Radial Basis Function Neural Network (RBF NN)
Classification Ability	Accuracy Rate	61.90	76.20
	Kappa Statistics	0.33	0.57
	Avg. precision	1.88	1.86
	Avg. recall	1.80	2.04
	F1-Score	1.83	1.94

To compare the performance of ANN models, various classification metrics were employed and are presented in Table 7. These measures aid in selecting the most suitable model for classification purposes. The accuracy rate achieved by MLP NN is 61.90 percent, whereas RBF NN has higher accuracy rate of 76.20 percent.

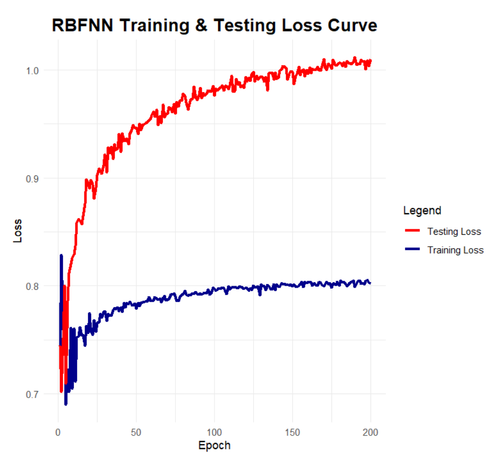

Fig. 6(a). Loss vs Epoch graph for of RBFNN

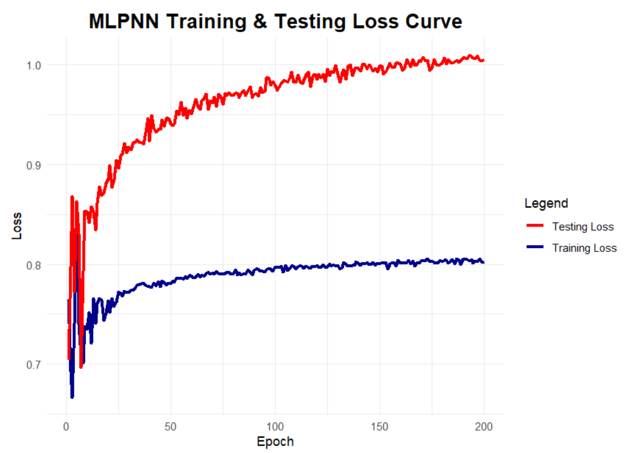

Fig. 6(b). Loss vs Epoch graph for of MLPNN

Furthermore, the kappa statistics for MLP NN and RBF NN are 0.33 and 0.57, respectively. And, the average precision values for MLP NN and RBF NN are 1.88 and 1.86, respectively, with average recall values of 1.80 for MLP NN and 2.04 for RBF NN. Also, value of F1-Score for MLP NN is 1.83 and for RBF NN is 1.94 respectively. Moreover,based on the Loss versus Epoch graphs for both the neural network models, the training dataset values show a consistent decrease relative to the testing dataset as the number of epochs increases. This trend indicates that the models are learning effectively and validates the reliability of the results. Consequently, across all evaluation measures, RBF NN demonstrates greater precision in classification compared to MLP NN as per the measures.

4. CONCLUSION

The RBF NN techniques outperformed the MLP NN method in classifying rice genotypes across different classes of yield, exhibiting higher values in classification ability measures as RBF NN has accuracy rate of 76.20 percent while accuracy rate of MLP NN is 61.90 percent.The important variable for classifying rice genotypes using RBF NN for yield is number of effective tillers per plant followed by plant height, grain length, ratio of grain length & grain width, grain width and number of days for 50percent flowering where as in case of MLP NN important variable for classification of rice genotypes is the variable plant height followed by the number of effective tillers per plant, grain length, and the grain length–to–grain width ratio.

REFERENCES

Chen, X., Y.Xun, W.Li, and J .Zhang (2010). Combining discriminant analysis and neuralnetworks for corn variety identification. *Computers and Electronics in Agriculture*. 71: S48-S53.

Galdon B.R., Mendez E.M., Havel J. and Diaz C. (2010). Cluster analysis and artificial neural networks multivariate classification of onion Varieties. *Journal of Agricultural and Food Chemistry*, 58(21): 11435-11440.

Gupta, S., Sharma., M., Jasrotia, N. and Mahajan, S. (2025). Comparative analysis for classification of rice genotypes using statistical and Artificial Neural Network models on the basis of maturity. *Model Assisted Statistics and Applications.* 19(4): 1-8.

Gupta S., Sharma M., Rizvi S.E.H., Salgotra R.K., Gupta S.K. (2023). Statistical and artificial neural network approaches for the classification of rice genotypes based on morphological characters. *In Special proceedings society of statistics computer and applications, 25th Annual Conference, 15–17 February 2023*, (pp. 131–137).

Harper, J.L., P.H. Lovell, and K.G. Moore (1970). The shapes and sizes of seeds. *Annu. Rev. Ecol. Syst*. 1: 327-356.

Jayas, D.S., J. Paliwal and N.S. Visen (2000). Multi-layer neural networks for image analysis of agricultural products. *J. Agricultural Engineering Research*. 77: 119-128.

Jasrotia, N. (2025). Scenario and Estimation of Walnut Production through Statistical and Artificial Neural Network models. *Ph.D thesis, Sher-e- Kashmir university of Agricultural Sciences and technology of Jammu, J&K.*

Pazoki, A.R. and Z. Pazoki (2011). Classification system for rain fed wheat grain cultivars using artificial neural network. *African J. Biotechnology*. 10(41): 8031-8038.

Kumar, R. and Verma, R. (2012). Classification algorithms for data mining: A survey.*Int. J. Innov. Eng. Technol, IJIET*.1(2): 7–14.

Shouche, S.P., R. Rastogi, S.G. Bhagwat, and J.K. Sainis (2001). Shape analysis of grains of Indian wheat varieties. *Computers and Electronics in Agriculture*. 33: 55-76.

Sharma M., Gupta R., Bhat A., Bhat M.I.J. and Sharma S. (2022). Price Model for Summer and Winter Tomato Crop through Discriminant function. *Journal of Community Mobilization and sustainable Development*. Vol.1 (SSI). pp:237-245.