Environment and Ecology 41 (1) : 63—72, January—March 2023
ISSN 0970-0420

Comparison of Classificatory Abilities of Artificial Neural Network and Support Vector Machines for Data Generated by Survey on Farmers

Halagundegowda G. R., Abhishek Singh

Received 2 September 2022, Accepted 3 November 2022, Published on 27 January 2023

ABSTRACT

Data collected from surveys are generated by processes which are difficult to ascertain and therefore can be modelled by black box approaches which are available in machine learning procedures. Machine learning has now become one of the most promising methods which are used for variety of purposes in agriculture. Studies have been conducted to compare various machine learning techniques,but in almost all of these studies the dependent variable is quantitative and SVM and ANN have been used as a substitute to regression techniques. In this study support vector machines (SVM) and artificial neural network (ANN) are used to classify farmers on the basis of their adoption behavior for drought coping mechanisms. Twelve socio economic variables are used to classify the farmers based on the data collected for farmers of Kolar Districts of Karnataka, India. It was found that the SVM performed better than the ANN in classifying the farmers.

Halagundegowda GR[1], Dr. Abhishek Singh*[2]
[1]PhD Scholar, [2]Associate Professor
Dept of Agricultural Engg, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi 221005, UP, India
Email: abhisheksingh@bhu.ac.in
*Corresponding author

Keywords Support vector machines, Artificial neural network, Adoption behavior, Kernel, Classification.

INTRODUCTION

Machine learning has now become one of the most promising methods which are used for variety of purposes in agriculture. Traditional classificatory techniques like discriminant analysis, cluster analysis have strict assumptions which need to be fulfilled in order to apply them successfully. Most of the times they are suitable for only linear processes, but now due to the availability of better computing facilities, machine learning has become the ultimate solution to classificatory problems. Most of the times these approached has been used in sciences like engineering, hydrology,medicine, but these approaches has not been used in case of data generated from surveys. In this paper two approaches namely support vector machine (SVM) and artificial neural network (ANN) has been compared in their classification ability for classifying farmers on the basis of certain variables into adopters and non adopters of drought coping mechanisms. Many traditional statistical methods like logisitic regression, discriminant analysis are available but their utility depends on the condition that certain assumptions are satisfied, these preconditions are not an hindrance for the application of machine learning methods like SVM and ANN.

Several studies have been conducted to compare the performances of SVM and ANN in diverse area like Khalil *et al.* (2013) concluded that SVM perform better in the testing phase as it minimizes the struc-

Table 1. Variables encoding summary.

Code	Variables	Measurement
Y	Adoption behavior	Y= 0 for Non-Adopters = 1 for Adopters
X_1	Age of the farmer	Number of years
X_2	Education of the farmer	Formal Years of Education
X_3	Household size	Number of family members
X_4	Farm size	Number of acre's
X_5	Farming experience	Number of years
X_6	Animal husbandry	Number of farm animals and poultry birds
X_7	Media exposure	Number of sources exposed frequently
X_8	Extension visits	Number of visits made to an research organizations
X_9	Crop diversification	Number of crops grown in that year
X_{10}	Income status	In rupees (Rs)
X_{11}	Worth of liquidating assets	In rupees (Rs)
X_{12}	Crop insurance got by the government	In rupees (Rs)

tural risk as compared to the empirical risk by ANN which indicates their ability to avoid overtraining and hence good generalization. Niu *et al.* (2019) compared three artificial intelligence algorithms (ANN, SVM and ELM) and found that the performance of SVM was better than ANN. Zhang *et al.* (2009) concluded that the SVM has better generalization ablility than ANN and recommended that SVM can be used to approximate SWAT model. Gao (2018) presented a contrary view in finding that the ANN performed better than the other competing machine learning tools. It was concluded that machine learning and nonparametric algorithms had limited effects on improving estimation; but ANN was relatively the best model in this study. Raczko (2017) recommended that more studies should be conducted to investigate optimal data selection techniques and found that ANN performed better than SVM or RF. Prasad *et al*. (2017) in their study found that SVM performed better than the ANN in LULC classification of high resolution landsat-8 satellite images. Dos Reis *et al.* (2018) found that machine learning algorithms, particularly the random forest (RF) and support vector machine (SVM) algorithms performed better than the ANN. Yao *et al.* (2010) also found that the errors from the SVM models are less than that of ANN and the possible reason could be that SVM utilizes structural risk minimization priniciple while the ANN uses the empirical risk minimization priniciple and therefore the SVM always tend to find the global minimum while ANN falls into a local optimal solution. Judson *et al.* (2008), concluded that ANN and SVM were always in the top performing set of methods .Kalantar *et al.* (2018) showed that LR and ANN models had better performances than the SVM model in terms of sensitivity to training samples. Lu *et al.* (2012) also concluded that RBF–SVM performed better than other methods in the classification ability. Dou *et al.* (2018) concluded that the SVM model with the radial basis kernel function produced the most accurate estimates and performed substantially better than the SVM models with the polynomial and sigmoid functions.Wang *et al.* (2016) also showed that SVR performed better than the ANN. Some of the studies have also been conducted in the field of agriculture Halagundegowda *et al.* (2017, 2018), Nagaraja *et al.* (2018) and Jhade and Singh (2019). But in almost all of these studies the dependent variable is quantitative and SVM and ANN have been used as a substitute to regression techniques.

MATERIALS AND METHODS

The current study utilizes both classification and prediction techniques. The household secondary data was used to fit the classificatory statistical models and the data were recorded on socio- economic characters of farmers of Kolar Districts of Karnataka, India. The data is mainly related to coping strategies implemented against drought by the farmers of this region.

Variables of interest

The objective was to classify the farmers in to adopters and non-adopters based on their socio-economic characteristics which utilises the following information given in Table 1.

Case processing summary

Table 2 shows that 120 cases were assigned to the training sample and 30 sample as testing sample, this is in conformity with the rules that 80% of data set as training sample and 20% of data set as testing sample. The training sample comprises the data records used for training in order to obtain a model. The testing sample is an independent set of data records used to track errors during training in order to prevent over-training. A numeric variable was used to assigns each case in the active dataset to the training and testing data set. Cases with a positive value on the variable are assigned to the training sample and cases with a value of 0, to the testing sample.

Artificial neural network (ANN) model

Neural networks are simulated networks with interconnected simple processing neurons which aim to mimic the function of the brain central nervous system. In 1943 for the first time the idea of the artificial neural network (ANN) was proposed but because of the lack of computing facilities they were not in much use until the back propagation algorithm was discovered in 1986. Neural networks are good at input and output relationship modeling even for noisy data. The greatest advantage of a neural network is its ability to model complex non linear relationship without a priori assumptions of the nature of the relationship. Apart from this artificial neural networks can also be used for classification problems. The ann model performs a nonlinear functional mapping from the past observations (Xt-1 , Xt-2,……….., Xt-p) to the future value Xt i.e.

$$X_t = f(X_{t-1}, X_{t-2}, \ldots\ldots\ldots, X_t\text{-}p, w) + e_t \qquad (2)$$

Where w is a vector of all parameters and f is a function determined by the network structure and connection weights.

Table 2. Case processing summary.

DATA set		N	Percentage
Sample	Training	120	80.0 %
	Testing	30	20.0 %

Training of the neural network is essential factor for the success of the neural networks. Among the several learning algorithms available, back propagation has been the most popular and most widely implemented learning algorithm of all neural networks paradigms. The important task of the ANN modeling to choose an appropriate number of hidden nodes, q, as well as the dimensions of the input vector p .

Support vector machine for classification

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. Linear SVMs can very easily be generalized to include nonlinear decision functions. The so-called kernel trick can accomplish this generalization.

In this article, binary class SVM by solving single optimization problem are used. The hyper-parameters of this model are estimated using very efficient random grid search technique.

Reliability measures for classification ability of model

Hit rate **:** Number of correct predictions divided by sample size. The hit rate for the model should be compared to the hit rate for the classification table

Table 3. Model summary. a. error computations are based on the testing sample.

Training	Cross entropy error		33.178
	Percent incorrect predictions		10.8%
	stopping rule used	1 consecutive step (s) with no decrease in error[a]	
	Training time		0:00:00.10
Testing	Cross entropy error		5.816
	Percent incorrect predictions		6.7%

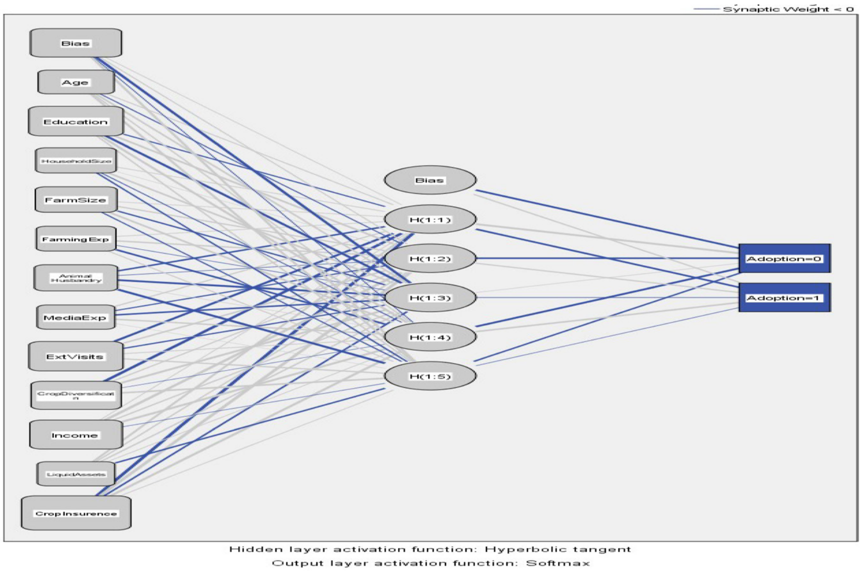

Fig. 1. Architecture of the network.

for the constant-only model.

***Sensitivity* :** Percent of correct predictions in the reference category (usually 1) of the dependent. It also refers to the ability of the model to classify an event correctly.

***Specificity* :** Percent of correct predictions in the given category (usually 0) of the dependent. It also refers to ability of the model to classify a non-event correctly.

***False positive rate* :** It is the proportion of predicted event responses that were observed as non-events.

***False negative rate* :** It is the proportion of predicted non-event responses that were observed as events.

Higher the sensitivity and specificity lower the false positive rate and false negative rate, better the classificatory ability.

Table 4. Area under the curve.

Adoption	Area
Non adopters	0.968
Adopters	0.968

RESULTS AND DISCUSSION

Artificial neural network model

A multilayered perceptron neural network was fitted to the data with the help of SPSS 22.0 statistical package, the number of hidden nodes varied from 2 to 10. Thus, different numbers of neural network models are tried before arriving at the final structure of the model. Out of all neural network structures a neural network model with 12 input nodes and 5 hidden nodes performed better than other competing

Table 5. Classification summary.

Sample	Observed	Predicted		
		Non adopters	Adopters	Percent correct
Training	Non adopters	43	7	86.0%
	Adopters	6	64	91.4%
	Overall percent	40.8%	59.2%	89.2%
Testing	Non adopters	11	1	91.6%
	Adopters	2	16	88.9%
	Overall percent	46.7%	53.3%	90.2%

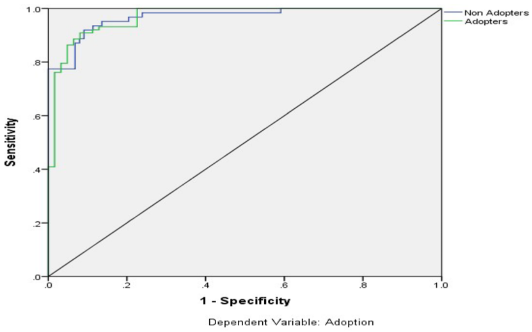

Fig. 2. ROC curve.

models in respect of out-of sample prediction and classification of adoption behavior of farmers.12 input variables are used as covariates in analysis and the standardized rescaling method is used for adjusting the covariates. Scale dependent variables and covariates are rescaled by default to improve network training. All rescaling is performed based on the training data, even if a testing sample is defined. The network has an input layer with 12 input nodes; the number of units in the input layer is the number of covariates. A single hidden layer with 5 hidden nodes and an output layer with 2 output nodes.

The hyperbolic tangent function used as activation function in hidden layers and it takes real valued arguments then transforms them to the range (–1, 1). The error is the cross entropy error because softmax activation function is applied to the output layer. It takes a vector of real valued arguments and transforms it to a vector whose elements fall in the range (0, 1) and sum to 1. Softmax is available only if all dependent variables are categorical.

Table 6. Model summary.

Model specifications	Value
Number of independents	12
SVM type	Classification type 2 (capacity= 2.40)
Kernel type	Sigmoid (Gamma= 0.096, coefficient= 0.06)
Number of SVs	97 (93 bounded)
Number of SVs (0)	48
Number of SVs (1)	49

The architecture of the network has been shown in the Fig.1. light color lines display weights greater than zero and the dark color lines show weight less than zero. The number of covariates and factor levels increases, the diagram becomes more difficult to interpret.

Table 3 displays information on the result of training and applying the final network to the testing sample. Cross entropy error is displayed because the output layer uses the softmax activation function. This is the error function that the network tries to minimize during training. Cross-entropy error will have a predicted value for each category, where each predicted value is the probability that the case belongs to the category.

In the above table the cross entropy error is 33.18 which is tolerable level and can continue the analysis for further steps also. The percentage of incorrect predictions is taken from the classification table and there is 10.8% of predictions which are miss match with the original observed samples. Here there is one step to allow before checking for a decrease in error. The estimation algorithm stopped because the maximum number of epochs was reached. Ideally, training should stop because the error has converged. The cross entropy error is 5.8 and 6.7% incorrect predictions for testing data. There is decline in both entropy error and incorrect predictions.

Fig. 2 Displays an ROC (Receiver Operating Characteristic) curve for each categorical dependent variable and also displays a table giving the area under each curve. For a given dependent variable, the ROC chart displays one curve for each category. If the dependent variable has two categories, then each curve treats the category at issue as the positive state versus the other category. If the dependent variable has more than two categories, then each curve treats the category at issue as the positive state versus the aggregate of all other categories.

The ROC curve gives you a visual display of the sensitivity and specificity in a single plot, which is much cleaner and more powerful than a series of tables. The chart shown here displays two curves, one for the category of non-adopters and another for

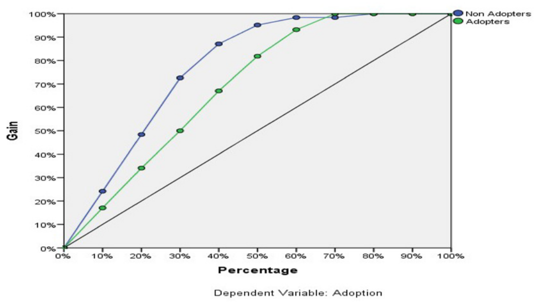

Fig. 3. Cumulative gain Chart.

the category of adopters. Since there are only two categories, the curves are symmetrical about a 45° line (not displayed) from the upper left corner of the chart to the lower right. Note that this chart is based on the combined training and testing samples.

Table 4 depicts that the area under the curve is a numerical summary of the ROC curve and the values in the table represent for each category, the probability that the predicted pseudo probability of being in that category is higher for a randomly chosen case in that category than for a randomly chosen case not in that, considering the current research for a randomly selected adopters and randomly selected non-adopters, there is a 0.968 probability that the model predicted pseudo probability of adoption will be higher for the adopters than for the non-adopters. While the area under the curve is a useful one statistic summary of the accuracy of the network, there is need to be able to choose a specific criterion by which farmers are classified. The predicted by observed chart provides a visual start on this process.

Fig. 3 displays a cumulative gains chart for each categorical dependent variable (Adopters and Non-Adopters). The display of one curve for each dependent variable category is the same as for ROC curves. The cumulative gains chart shows the percentage of the overall number of cases in a given category "gained" by targeting a percentage of the total number of cases. In the above chart the first point on the curve for the Non-Adopters category is at (15%, 25%), meaning that if you score a dataset with the network and sort all of the cases by predicted pseudo probability of No (Non-Adopters), you would expect the top 15% to contain approximately 25% of all of the cases that actually take the category No (Non-Adopters). Likewise, the top 35% would contain approximately 50% of the Non-Adopters; the top 50% of cases would contain 70% of Non-Adopters and so on. If you select 100% of the scored dataset, you obtain all of the Non-Adopters in the dataset.

The diagonal line is the baseline curve, if we select 10% of the cases from the scored dataset at random; we would expect to "gain" approximately 10% of all of the cases that actually take the category Yes (Adopters). The farther above the baseline a curve lies, the greater the gain. In the above chart if we select 10% of the cases from the scored dataset at random; we would expect to "gain" approximately 15% of all of the cases that actually take the category Yes (Adopters) and 25% of all of the cases that actually take the category No (Non-Adopters). Likewise if we select 20% of the cases from the scored dataset at random; we would expect to "gain" approximately 35% of all of the cases that actually take as Adopters and 50% of all of the cases that actually take as Non-Adopters. We can use the cumulative gains chart to help choose a classification cut off by choosing a percentage that corresponds to a desirable gain and then mapping that percentage to the appropriate cut off value.

The "desirable" gain depends on the cost of Type I and Type II errors. That is, the cost of classifying a Non-Adopters as a Adopter (Type I) and the cost of classifying a Adopter as a Non-Adopter (Type II). If Adoption behavior is the primary concern, then we want to lower our Type I error on the cumulative gains chart, this might correspond to rejecting the farmers in the top 40% of pseudo predicted probability of No, which captures nearly 90% of the possible Non-Adopters but removes nearly half of our farmers' pool. If growing our farmer's base is the priority, then we want to lower our Type II error. In the above chart, this might correspond to rejecting the top 15%, which captures 25% of the Non-Adopters and leaves most of our farmers' pool intact. Usually both are major concerns; hence need to choose a decision rule for classifying farmers that gives the best mix of sensitivity and specificity.

Fig. 4 shows that, for categorical dependent

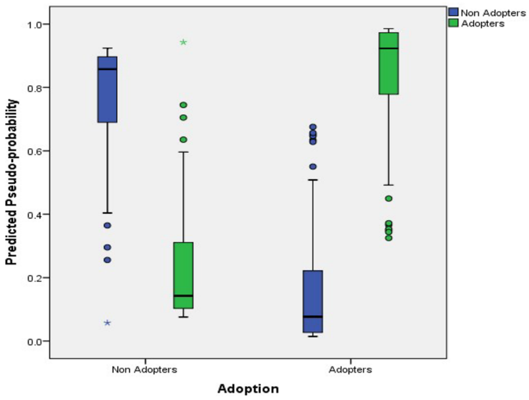

Fig. 4. Predicted by observed chart.

variables the predicted by observed chart displays clustered boxplots of predicted pseudo probabilities for the combined training and testing samples. The x axis corresponds to the observed response categories and the legend corresponds to predicted categories.

The leftmost boxplot shows for cases that have observed category No, the predicted pseudo probability of category No. The portion of the boxplot above the 0.5 mark on the y axis represents correct predictions shown in the classification table. The portion below the 0.5 marks represent incorrect predictions. Remember from the classification table that the network is very good at predicting cases with the No category using the 0.5 cut off, so only a portion of the lower whisker and some outlying cases are misclassified.

The next boxplot to the right shows for cases that have observed category No, the predicted pseudo probability of category Yes. Since there are only two categories in the target variable, the first two boxplots are symmetrical about the horizontal line at 0.5. The third boxplot shows, for cases that have observed category Yes, the predicted pseudo probability of category No. It and the last boxplot are symmetrical about the horizontal line at 0.5.

The last boxplot shows, for cases that have observed category Yes, the predicted pseudo-probability of category Yes. The portion of the boxplot above the 0.5 mark on the y axis represents correct predictions shown in the classification table. The portion below the 0.5 mark represents incorrect predictions. Remember from the classification table that the network predicts slightly more than half of the cases with the Yes category using the 0.5 cut off, so a good portion of the box is misclassified.

The classification table shows the practical results of using the network. For each case, the predicted response is Yes if that cases' predicted pseudo probability is greater than 0.5. Unfortunately, the single cut off value gives you a very limited view of the predictive ability of the network, so it is not necessarily very useful for comparing competing networks. Instead look at the ROC curve.

Table 5 shows that the cells on the diagonal of the cross classification of cases are correct predictions for each sample. The cells off the diagonal of the cross classification of cases are incorrect predictions of the cases used to create the model, 64 of the 70 farmers who previously adopted the drought coping strategies are classified correctly. 43 of the 50 non-adopters are classified correctly. Overall, 89.2% of the training cases are classified correctly, corresponding to the 10.8% incorrect shown in the model summary table. A better model should correctly identify a higher percentage of the cases.

Classifications based upon the cases used to create the model tend to be too "optimistic" in the sense that their classification rate is inflated. The testing sample helps to validate the model, here 90.2% of these cases were correctly classified by the model. This suggests that overall our model is in fact correct and efficient in prediction and classification.

Table 7. Classification summary.

Sample	Observed	Predicted Non - adopters	Adopters	Percent correct
Training	Non adopters	42	8	84.0%
	Adopters	8	62	88.5%
	Overall percent			86.2%
Testing	Non adopters	11	1	91.6%
	Adopters	1	17	94.4%
	Overall percent			93.0%

Support vector machine model

Support Vector Machines (SVMs) is a Generalized Portrait classification algorithm based on statistical learning theory and developed to perform binary classification problem initially. In this article, we have employed binary class SVM by solving single optimization problem. The hyper-parameters of this model are estimated using very efficient random grid search technique. Sigmoid kernel method of SVM was fitted to the data with the help of STATISTICA_8.0.360 statistical package, although many procedures are available in traditional statistical classification, the usefulness depends on assumptions and circumstances.

Table 6 explains the detail about the model summary and specifications of the SVM model, including the number of support vectors and their types, the kernels and their parameters. The model was constructed to classify the binomial dependent variable by including 12 predictors and there are two major classifications types in SVM such as classification type 1 and classification type 2, classification type1 means SVM applies for linearly separable data and classification type 2 means SVM applies for linearly non separable data. The capacity value indicates the trade off between two boundaries, which is the hyper parameter of the predictive model, the current research take up with classification type 2 with capacity C= 2.40.

In classification problems, only two hyper-parameters are needed to be defined by user i.e. the trade-off between model capacity and training error represented by ((capacity) and the kernel parameter γ (Gamma). These hyper parameters are directly coded with real values within a given search space to randomly generate M number of initial particles of swarm set *S. search* space of hyper-parameters are respectively restricted to ranges of $[C_{min}, C_{max}]$, $[\gamma_{min}, \gamma_{max}]$ are randomly generated. The entire operations has been system inbuilt algorithms in STATISTICA 8 version software. The current research has identified the capacity parameter, by randomly grid search method and the model provides saturated result at the value C=2.40. For finding the optimal value of hyper parameters traditionally various optimization

Table 8. Comparison of Models based on Classification and Predicition ability.

Measures (%)	Classification ability ANN MLP	SVM Sigm
Hit rate	90.00	93.33
Sensitivity	88.89	94.44
Specificity	90.66	91.67
False positive rate	11.11	5.56
False negative rate	8.33	8.33

techniques such as particle swarm optimization, genetic algorithms, ant colony optimization techniques, simulated annealing algorithms etc. but the current research have taken random grid search algorithm by specifying the range of values between minimum and maximum with particle incremental value as an interval.

Further, several kernel functions $k\ (xi,\ xj)$are available in the literature, like Polynomial function, Sigmoid Kernel, Gaussian kernel and Radial basis Function (RBF). The current research has taken up with Sigmoid kernel and having hyper parameter gamma=0.096, which has achieved by selecting the parameters range $[\gamma_{min}, \gamma_{max}$. The γ_{min}, =0 and γ_{max} =10 with the incremental value =0.2 by providing this information to the random grid search would end up with the saturated result of gamma γ=0.096 and coefficient value is mainly depends on how exactly the position and orientation of hyper plane and which is depends on the weights which we assign to the input vectors. It's ranged from 0 to 3 ranges with the incremental value 0.02 wind up with the saturated final result 0.06. Set the iteration number (t) from 1 to maximum number of iterations and evaluate inertia weight w (t) generation by generation according to the model equation. The current research has taken i=1000 iterations for tuning the model.The current research has 97 overall support vectors and having 93 bounded support vectors and coming to category wise, there are 48 support vectors are in the side of non-adopters category and around 49 support vectors are in the side of adopters category. The result indicates there is sufficient number of vectors is at boundaries which are making the hyper plane for effectively classifying the cases into respective classes.

Table 7 shows that the cells on the diagonal of the cross classification of cases are correct predictions for each sample. The cells off the diagonal of the cross classification of cases are incorrect predictions of the cases used to create the model, 62 of the 70 farmers who previously adopted the drought coping strategies are classified correctly. 42 of the 50 non-adopters are classified correctly. Overall, 86.2% of the training cases are classified correctly, corresponding to the 13.8% incorrect shown in the model summary table. A better model should correctly identify a higher percentage of the cases.

Classifications based upon the cases used to create the model tend to be too "optimistic" in the sense that their classification rate is inflated. The testing sample helps to validate the model; here 93% of these cases were correctly classified by the model. This suggests that overall our model is in fact correct and efficient in prediction and classification.

The Table 8 provides the result of various classification measures such as hit rate, sensitivity, specificity, false positive rate, false negative rate and % accuracy for testing samples for both the machine learning models. The models having high hit rate, sensitivity, specificity, % accuracy and whereas the models with low false positive rate and low false negative rate were considered as best models for classification purpose. The SVM model has hit rate of 93.33 as compared to 90.00 of ANN, sensitivity of 94.44 as compared to 88.89 of ANN, specificity of 91.67 as compared to 90.66 of ANN indicating that in this case SVM performed relatively better than the ANN. Also the false positive rate of SVM was 5.56 which was lesser than that of the ANN (11.11). So for the data generated a by a process which was clearly not comprehensible and therefore Black box in nature the SVM performed better than the ANN in classificatory abilities.

CONCLUSION

Classification is a data mining technique used to predict group membership for data instances in order to classify the objects based on their features and characteristics, it is one of the most important and primitive activities of social research. By consideration of the objectives of the present research study, the data recorded on adoption behavior of drought coping strategies across various socio-economic characters of farmers of Kolar Districts of Karnataka and analysed using statistical techniques such as Artificial Neural Network and Support Vector Machines were used in the present investigation.

Adoption of drought coping strategies is considered as categorical response variable and it's a mental perception process, in which farmers deciding whether to adopt the strategies or not. This process affected by various socio-economic, agro ecological, institutional and resource based factors as explanatory variables such as age, education status, house hold size, farm size, farming experience, animal husbandry, media exposure, extension visits, crop diversification, income status, worth of liquidating assets and crop insurance.

REFERENCES

Dos Reis AA, Carvalho MC, de Mello JM, Ferraz Filho AC, Acerbi FW (2018) Spatial prediction of basal area and volume in Eucalyptus stands using Landsat TM data : An assessment of prediction methods. *New Zealand J For Sci* 48 (1) : In press.

Dou X, Yang Y (2018) Comprehensive evaluation of machine learning techniques for estimating the responses of carbon fluxes to climatic forces in different terrestrial ecosystems. *Atmosphere* 9 (3) : 83.

Gao Y, Lu D, Li G, Wang G, Chen Q, Liu L, Li D (2018) Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. *Remote Sens* 10 : 627.

Halagundegowda GR, Singh A (2018) Multilayer perceptron method of artificial neural network for classification of farmers based on adoption of drought coping mechanisms. *Int J Pure Appl Biosci* 6 (2) : 1408—1414.

Halagundegowda GR, Singh A, Meenakshi HK (2017) Discriminant analysis for prediction and classification of farmers based on adoption of drought coping mechanisms. *Agric Update* 12 (Techsear -3) : 635—340.

Jhade S, Singh A (2019) Evaluation of the performance of supervised classification algorithms in image classification. *Int J Curr Microbiol Appl Sci* 8 (10) : In press.

Judson R, Elloumi F, Woodrow RW, LZ Shah I (2008) A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. *BMC Bioinformatics* 9 : 241.

Kalantar B, Pradhan B, Amir Naghibi S, Motevalli A, Mansor S

(2018) Assessment of the effects of training data selection on the landslide susceptibility mapping : A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN) Geomatics, *Natural Hazards Risk* 9 (1) : 49—69.

Khalil A, Almasri MN, McKee M, Kaluarachchi JJ (2005) Applicability of statistical learning algorithms in groundwater quality modeling. *Water Resour Res* 41(5)W05010: 1—16.

Lu H, Jiang W, Ghiassi M, Lee S, Nitin M (2012) Classification of camellia (theaceae) species using leaf architecture variations and pattern recognition techniques. *Plos One* 7 (1) : e29704.

Nagaraja MS, Singh A (2018) Use of ordinal logistic regression and multiclass discriminant model for classification of genotypes for maturity of little millet. *Int J Pure Appl Biosci* 6 (6) : 248—258.

Niu WJ, Feng ZK, Feng BF, Min YW, Cheng CT, Zhou JZ (2019) Comparison of multiple linear regression, artificial neural network, extreme learning machine and support vector machine in deriving operation rule of hydropower reservoir. *Water* (Switzerland) 11(1) : 88.

Prasad SVS, Savithri TS, Krishna IVM (2017) Comparison of accuracy measures for RS image classification using SVM and ANN classifiers. *Int J Electric Computer Engg* 7(3) : 1180—1187.

Raczko E, Zagajewski B (2017) Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Europ J Remote Sensing* 50 (1) : 144—154.

Wang L, Zhou X, Zhu X, Dong Z, Guo W (2016) Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J* 4 (3) : 212—219.

Yao BZ, Yang CY, Yao JB, Sun J (2010) Tunnel surrounding rock displacement prediction using support vector machine. *Int J Computational Intelligence Syst* 3 (6) : 843—852.

Zhang X, Srinivasan R, Van Liew M (2009) Approximating SWAT model using artificial neural network and support vector machine. *J Am Water Resources Association* 45 (2) : 460—474.