PLANT SCIENCE TODAY
Vol 13(sp1): 1-9
https://doi.org/10.14719/pst.11723

eISSN 2348-1900

RESEARCH ARTICLE

Application of artificial neural networks and random forest algorithms for forecasting marketing security among agarwood growers in Assam

Rituraj Boruah[1], Hemanga Kumar Kalita[1], Rabijita Buragohain[2], Pallabi Das[1*], Utpal Barman[1], Nivedita Deka[3], Jayanta Kumar Dutta[1], Debajit Borah[4] & Chittaranjan Deka[5]

[1]Department of Extension Education, College of Agriculture, Assam Agricultural University, Jorhat 785 013, Assam, India
[2]Department of Agricultural Statistics, College of Agriculture, Assam Agricultural University, Jorhat 785 013, Assam, India
[3]Department of Agricultural Economics and Farm Management, College of Agriculture, Assam Agricultural University, Jorhat 785 013, Assam, India
[4]Department of Extension Education, College of Sericulture, Assam Agricultural University, Jorhat 785 013, Assam, India
[5]Extension Education Institute (NE Region), Assam Agricultural University, Khanapara 781 022, Assam, India

*Correspondence email - pallabi.das@aau.ac.in

Received: 10 September 2025; Accepted: 01 February 2026; Available online: Version 1.0: 24 March 2026

Cite this article: Rituraj B, Hemanga KK, Rabijita B, Pallabi D, Utpal B, Nivedita D, Jayanta KD, Debajit B, Chittaranjan D. Application of artificial neural networks and random forest algorithms for forecasting marketing security among agarwood growers in Assam. Plant Science Today. 2026; 13(sp1): 1-9. https://doi.org/10.14719/pst.11723

Abstract

Benefits earned from a single agar (*Aquilaria malaccensis*) plant make it unique and popular among growers in Assam. The present study was conducted in 2024 to assess the marketing security of agar growers using different classifiers viz., artificial neural network (ANN) and random forest. A total of 420 agar growers were selected following the snowball sampling technique. The findings indicated that 45.95 % of the agar growers had a medium level of marketing security. Results of multiple regression analysis showed that 91.55 % variation in the marketing security of agar growers was explained by seven independent variables, namely age, family occupation, annual income, experience in agar cultivation, indebtedness, economic motivation and access to extension services. As per stepwise regression analysis, eight variables viz., age, family occupation, annual income, experience in agar cultivation, indebtedness, economic motivation, access to extension services and risk orientation had a high influence on marketing security, with an R^2 value of 96.36 %. The multilayer perceptron neural network (MPN) model indicated that five independent variables viz., age, annual income, economic motivation, access to extension services and risk orientation, generated the greatest number of activated neurons influencing the marketing security of agar growers through the hidden layer H (1:1) and H (2:3). The normalized graph depicted that annual income had the maximum influence on the marketing security of agar growers.The random forest classifier established that age was the most influential variable.

Keywords: marketing channel; marketing security; multilayer perceptron; neural network; random forest

Introduction

Aquilaria malaccensis is one of the economically important tree species distributed in North-east India, which produces heartwood known as agarwood—the most expensive wood in the world (1). It is presently being cultivated on a large scale, mainly in the Upper Assam region, in association with other useful plants, for its high commercial value. Currently, many home gardeners have started growing agar for greater returns and the species has become an important plantation crop of Upper Assam, especially in Golaghat and parts of Jorhat and Hojai district of Assam. These districts have now become important loci for small-scale agarwood chip preparation and oil extraction (2). Growers generally sell standing trees to local traders, who, in turn, sell the wood and woodchips to agar oil extraction industries. The cultivation of agar tree species has grown dramatically over the last 10–15 years. This development is driven by the desire to produce sustainable agarwood to meet the high market demand for oil. Prior to the 1991 export ban on wood and wood products, Mumbai served as the main exporting centre to Middle Eastern countries. Agarwood harvested from the North-Eastern states—predominantly Assam was taken to Hojai, where it was processed into chips, dust and oil (3, 4). Recently, the Assam government has relaxed the restrictions on the cultivation of the scented agar tree. The government has promoted agar tree cultivation and removed the hindrances to agarwood production. It has allowed plantation and sale of agar up to five hectares (35 bighas) of land and approved the growing and harvesting of agar on non-forest lands. The government also plans to set up an International Trade Centre for agar in Golaghat in the coming years. The number of agar trees found in the non-forest zones of Assam is estimated at 1.433 million, with 91 % of them distributed across four districts: Jorhat, Golaghat, Sivasagar and Hojai (5).

Benefits earned from a single agar plant make it unique and popular among growers. If properly cultivated and marketed, this wonder tree can change the rural economy of Assam-particularly the upper, central and southern parts of Assam. The value of an agar plant is determined by the growers and the buyers based on physical and manual verification, which is a primitive method. Moreover, the marketing chain of this product is not scientific or organized. Many times, middlemen take a considerable amount of profit. Hence, the present research was conducted to examine the marketing security of agar *(A. malaccensis)* growers in Assam.

Materials and Methods

The research study was conducted using an ex-post facto research design in Jorhat, Golaghat, Sivasagar and Hojai districts of Assam, considering a total of 420 agar growers by following snowball sampling technique. To collect the data for the present study, respondents were selected from six villages in Jorhat district under three agar-growing areas, namely Titabor, Selenghat and Pokamura. Seven villages in Sivasagar district were considered from the agar-growing area i.e., Namti, whereas eight villages in Golaghat district were taken under the Naharani area and seven villages in Hojai district were selected from the agar-growing areas i.e., Modertoli and Udali. Respondents were agar growers from the selected districts. However, it is evident that the use of snowball sampling for a study aiming for predictive modeling introduces selection bias and limits the generalization of the findings.

The marketing security of agar growers was analyzed using different dimensions, such as change in income level, trustworthiness, payment received in advance due to agar cultivation, expertise in price fixation of agar products, perceived fair price received for agar products, frequency of visits by agar buyers, nature of monetary transactions, availability of alternative marketing sources, access to market information, availability of transport, storage facilities and structured markets. The responses were obtained on a two-point continuum namely “Yes” and “No” with assigned scores of 1 and 0 respectively. The marketing security score of a respondent was calculated by adding up the scores obtained on all statements. A higher score indicated that the respondent had a higher level of marketing security.

For analysis, various statistical tools viz., quartiles along with multiple regression, stepwise regression, artificial neural network (ANN) and random forest classifiers were used.

Quartile

Quartiles were used to partition the obtained marketing security scores of agar growers, which split the dataset into four equal parts of 25 % each, with three quartiles, i.e., Q1, Q2 and Q3. Based on the quartiles of the scores obtained, the respondents were classified into three categories as follows:

It is used to divide the dataset scores into three categories i.e., low, medium and high. The values of observations in a dataset, when arranged in an ordered sequence, can be divided into four equal parts or quarters using three quartiles, namely Q1, Q2 and Q3. The first quartile (Q1) divides a distribution in such a way that 25 % (= n/4) of the observations have values less than Q1 and 75 % (= 3n/4) have values greater than Q1. The second quartile (Q2) has an equal number of observations above and below it and is the same as the median. The third quartile (Q3) divides the dataset such that 75 % of the observations have values less than Q3 and 25 % have values greater than Q3.

Multiple regression analysis

To examine the combined effect of the independent variables on the dependent variable, a regression analysis was conducted. In this case, the proposed model assumes that all 18 predictors (independent variables) have a combined influence on the dependent variable. More specifically, variables with significant *p*-values in the regression analysis are considered good predictors for the model.

Stepwise multiple regression

Stepwise regression is a variation of multiple regression that provides a means of identifying independent variables that give the best possible prediction with the fewest independent variables. It allows the user to solve a sequence of one or more multiple linear regression problems through the stepwise application of the least square method. At each step in the analysis, a variable is added or removed, resulting in the greatest reduction in the error sum of squares. An earlier study reported that the method of stepwise multiple regression analysis is to insert variables in turn until the regression equation is satisfactory (6). The order of insertion is determined by using the partial correlation coefficient as a measure of the importance of variables not yet included in the equation.

Artificial neural networks (ANNs)

ANNs, often referred to as neural networks (NNs), are computer systems that draw inspiration from the biological neural networks seen in the human brain. ANN, or simply NN, consists of an input layer of neurons (or nodes/units), one or two (or even three) hidden layers of neurons and a final output layer neuron. Each connection is associated with a numerical value called a weight. The output, hi, of neuron ‘i’ in the hidden layer is,

$$hi = \sigma\left(\sum N_{j=1} V_{ij} X_j + T_i^{hid}\right)$$

where, σ = activation (or transfer) function

N = number of input neurons

V_{ij} = weights

Xj = inputs to the input neurons

T_i^{hid} = threshold terms of the hidden neurons

The purpose of the activation function is, besides introducing nonlinearity into the NN, to constrain the value of a neuron so that the network is not destabilized by divergent neurons. It has been depicted that an NN constructed in this manner can approximate any computable function to arbitrary precision. Numbers given to the input neurons are considered independent variables, while those returned from the output neurons are taken as dependent variables for the function being approximated by the NN. Inputs to and outputs from an NN can be binary (such as yes or no) or even symbolic (such as green, red, etc.) when the data are appropriately encoded. This feature measures a wide range of applicability to NNs.

Random forest

A random forest is a statistical algorithm used for both classification and regression. It is typically built from a collection of classification and regression trees. To create each tree, a random sample from the training dataset is chosen and used at each partition. Additionally, a random sample of predictors is selected (7). A constraint associated with decision trees is their tendency to become unstable in prediction upon the introduction of new datasets, which can lead to over fitting. This instability can be addressed by employing a combination of decision trees, each derived from an independent, random sample. Random forests' primary benefit is their ability to handle vast feature sets with ease and their effectiveness. As a result, the trees have the same distribution because each tree depends on a randomly selected vector that is sampled independently and identically. A random forest classifier is therefore a classifier composed of a collection of tree-like classifiers h (x, q_k), k = 1 . . . B, where the (q_k) are independent and identically distributed. Each of these trees votes for the popular class and the majority votes are used to classify. In the case of regression, the tree predictor h (x, q_k) takes on numerical values.

The trees generated in this manner are identically distributed and independent; the expectation of each tree is similar with any other tree. To improve the algorithm's performance, the aim is to reduce the variance. The B identically independent random variables, each with a variance s^2, has variance $1/B\sigma^2$. Assuming the variables as identical but not independent, then the trees are correlated and the pair wise correlation between them affects the variance and is shown as:

$$p\sigma^2 + \frac{1-p}{B}\sigma^2$$

As B increases, the last term in the equation vanishes and therefore the variance reduction is left to be a function of the correlation between pairs of trees. This reduction in correlation is achieved by considering random subsets of the features used in each tree. Therefore, although a subset of the observations is used for each tree, a subset of the covariates is also considered, following the general rule m £ p, where m denotes the number of independent variables chosen and p is the total number of variables.

a) Variable importance: The variable relevance quantifies the extent to which an increase in knowledge or the Gini coefficient index reduces the variance or the amount of impurity in each decision tree. A list of significant factors is produced by calculating the mean drop in impurity, which adds together the variables' average Gini index drop. It is expressed as follows:

$$V_{imp}(x_i) = \frac{1}{n_{trees}}[1 - \sum_{j=1}^{n_{tree}} GI(i)^j]$$

The model was trained using SPSS's default scaled conjugate gradient (SCG) optimization with early stopping (training stopped after one consecutive iteration without reduction in error). The sum-of-squares loss function was used. A specific learning rate or number of epochs is not reported because SCG does not use a fixed learning rate schedule and training stops automatically when the error no longer decreases. The final NN model selected by SPSS used a two-hidden-layer architecture. The first hidden layer contained 3 neurons and the second hidden layer contained 2 neurons, chosen automatically by SPSS based on model-fit heuristics. The hidden layers used the hyperbolic tangent activation function, while the output layer used an identity (linear) function suitable for continuous outcomes. The random forest model was trained using 500 trees (B=500). At each split, the algorithm randomly selected 3 predictor variables (m=3) to determine the best split. The minimum terminal node size was fixed at 5 observations, following the default regression setting in the random forest package.

Results and Discussion

Multiple regression analysis

From multiple regression analysis shown in Table 1, it was observed that 91.55 % of the variation in the marketing security of agar growers was explained by the independent variables selected for the present study. Out of the 18 independent variables, 7 variables significantly contributed to marketing security. Age, annual income, experience in agar cultivation, indebtedness, economic motivation and access to extension services were found to contribute significantly to the marketing security of agar growers at the 0.01 level of probability, while family occupation was found to be significant at the 0.05 level of probability. Thus, it can be stated that the 7 independent variables, namely age, family occupation, annual income, experience in agar cultivation, indebtedness, economic

Table 1. Contribution of independent variables on marketing security of agarwood growers

Sl. No.	Independent variables	Standardized coefficient (β)	Standard error	*p*- value
1	Age	0.079937	0.010312	7.55E-14**
2	Education	0.009222	0.027702	0.73939
3	Caste	-0.18912	0.09689	0.05165
4	Family size	-0.0148	0.029795	0.61973
5	Family occupation	0.135458	0.053673	0.011996*
6	Annual income	2.61E-05	2.57E-06	8.78E-22**
7	Experience in agar cultivation	0.02879	0.009944	0.003998**
8	Operational land holding	-0.02123	0.023603	0.368992
9	Indebtedness	-0.27234	0.098009	0.005714**
10	Economic motivation	0.123445	0.023756	3.24E-07**
11	Social participation	0.042777	0.0582	0.462774
12	Extension participation	0.053352	0.032109	0.097371
13	Mass media use	-0.04074	0.033483	0.224427
14	Access to extension services	0.14558	0.044136	0.001059**
15	Value orientation	0.03932	0.026539	0.139233
16	Risk orientation	0.05856	0.034204	0.087654
17	Cosmopoliteness	-0.01164	0.024372	0.633255
18	Planning ability	0.01882	0.038021	0.620877

R^2 = 91.55 %

*Significant at 0.05 level of probability

**Significant at 0.01 level of probability motivation and access to extension services, were good predictors of the marketing security of agar growers in Assam. These findings are in line with earlier investigations (8, 9).

Stepwise regression analysis

A stepwise selection strategy was adopted to select the most influential variables from 18 potential independent variables to check the influence on the dependent variable, i.e., marketing security of agar growers (Y). The selection process started with all potential independent variables in the model. It was found that in the first step, age (0.00<0.05) was the most influential variable on Y and was significant at the pre-assigned significance level of 0.05. Hence, in the next step, from the remaining 17 variables, one non-significant variable with the highest regression coefficient was dropped. It was seen that the accuracy of the model improved in the 3rd step, providing another most significant variable, annual income (0.00<0.05). The same drop-down process continued and the backward stepwise selection strategy established that at the 5th, 7th, 11th, 13th and 17th steps, economic motivation, access to extension services, experience in agar cultivation, indebtedness and risk orientation were found to be highly significant with *p*-values of 0.00, 0.00, 0.01, 0.01, 0.006 and 0.043 respectively. The model utilized a total of 19 steps to check all 18 variables, which showed that the remaining variables were non-significant and had less influence on the marketing security of agar growers. Thus, only these eight significant variables i.e., age, family occupation, annual income, economic motivation, access to extension services, experience in agar cultivation, indebtedness and risk orientation have highly influenced the marketing security of agar growers, with model efficiency of R^2 and adjusted R^2 values of 96.36 % and 91.17 % respectively. Similar results were reported in earlier studies (10, 11).

Stepwise regression model

From Table 2 and Fig. 1 & 2 the *p*-value was found to be less than 0.05; therefore, it was statistically significant at the 5 % level of significance, which implied that at least one of the coefficients of the explanatory variables was different from mean. The stepwise regression analysis confirmed that there were eight significant variables out of 18 independent variables viz., age (X1), family occupation (X5), annual income (X6), experience in agar cultivation (X7), indebtedness (X9), economic motivation (X10), access to extension services (X14) and risk orientation (X16), which had a high influence on the marketing security of agar growers. Thus, eight significant variables explained 91.36 % (R^2 value) of the dependent variable, i.e., marketing security of agar growers. The fitted model became:

Y= 6.535 + (0.078) X1 + (0.136) X5 + (2.62E-05) X6 + (0.018) X7 + (-0.267) X9 + (0.124) X10 + (0.149) X14 + (0.068) X16

i.e., marketing security of agar growers = intercept + (0.078) age + (0.136) family occupation + (2.62E-05) annual income + (0.018)

Table 2. Stepwise regression analysis of marketing security (Y) in relation to 18 causal variables (X1–X18)

	Anova test		Stepwise regression model					
	p- value	significance	Significant variable	Regression coefficient (β_1)	Standard error of (β_1)	*p*-value	Intercept (β_0)	Standard error of (β_0)
Model- II	7.1E-212	Yes	Age	0.078	0.010	5.36E-14	6.535	0.372
			Family occupation	0.136	0.052	0.010146		
			Annual income (Rs.)	2.62E-05	2.54E-06	2.66E-22		
			Experience in agar cultivation	0.018	0.007	0.01066		
			Indebtedness	-0.267	0.097	0.006454		
			Economic motivation	0.124	0.023	1.94E-07		
			Access to extension services	0.149	0.043	0.000602		
			Risk orientation	0.068	0.033	0.043781		

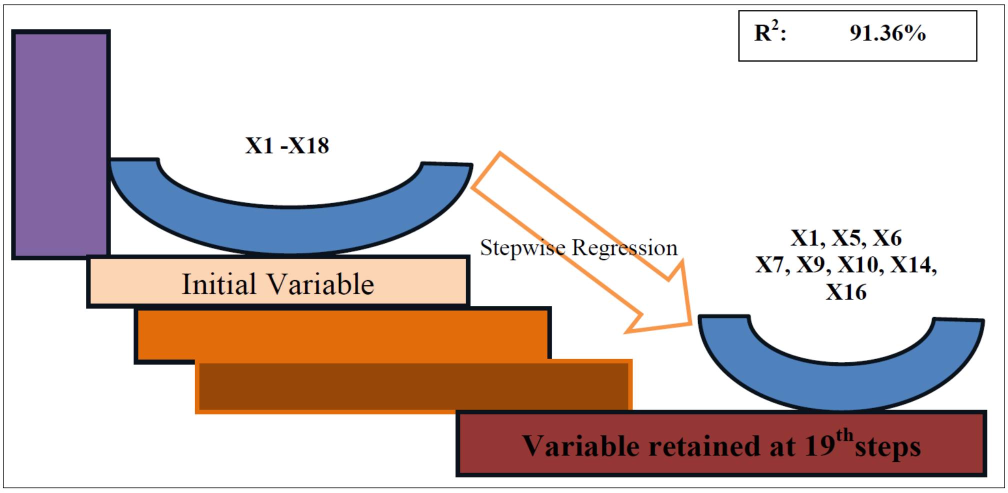

Fig. 1. Diagrammatic representation of significant variables in stepwise regression analysis of marketing security score (Y) against 18 independent variables.

Independent variables: X1 = Age; X2 = Education; X3 = Caste; X4 = Family size; X5 = Family occupation; X6 = Annual income; X7 = Experience in agar cultivation; X8 = Operational landholding; X9 = Indebtedness; X10 = Economic motivation; X11 = Social participation; X12 = Extension participation; X13 = Mass media use; X14 = Access to extension services; X15 = Value orientation; X16 = Risk orientation; X17 = Cosmopoliteness; X18 = Planning ability.

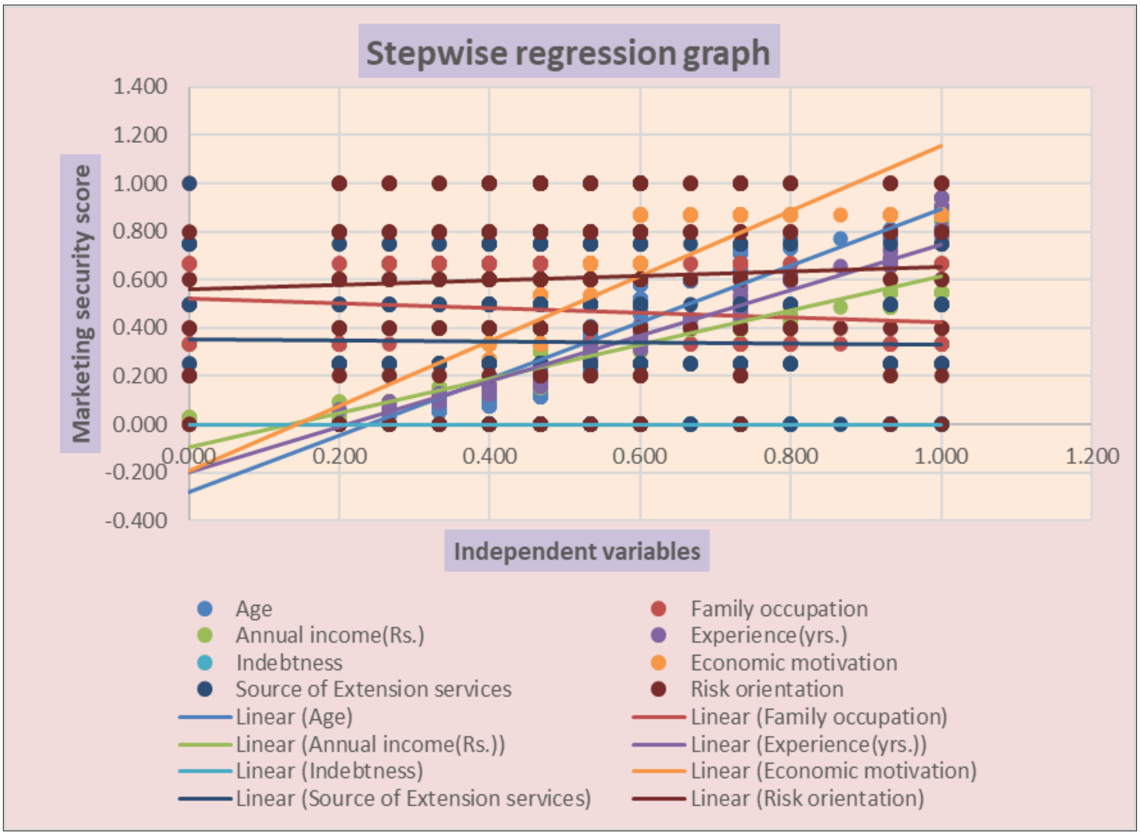

Fig. 2. Stepwise regression graph showing marketing security of agarwood growers (Y) against 18 independent variables.

experience in agar cultivation + (-0.267) indebtedness + (0.124) economic motivation + (0.149) access to extension services + (0.068) risk orientation

The above stepwise regression equation established that an increase in age, family occupation, annual income, experience in agar cultivation, indebtedness, economic motivation, access to extension services and risk orientation showed a direct relationship with the dependent variable i.e., marketing security of agar growers (Y). This implies that an increase in these variables resulted in an increase in the marketing security of agar growers. On the contrary, indebtedness showed an inverse relationship with the marketing security of agar growers (Y) indicating that an increase in this independent variable led to a decrease in the dependent variable.

Artificial neural network

A multilayer perceptron neural network (MPN) model was formed, with the dependent variable (Y) in the output layer and the input layer consisting of the predictors estimated through the stepwise regression process (12). The blue lines passing through the hidden layers (acting as moderators which are explained with values in the parameter estimates table; a positive value depicts a positive influence on the output) depict the interrelationships among the output and input layers.

Each unit is connected to all units at the next level and each unit receives input from all units at the previous level. Each connection has a specific weight, which illustrates the effect of one unit on the response of the unit to the next level. The output of a multilayer perceptron depends on the input and the connection strengths among the units. When information is fed into the multilayer perceptron through neurons at the input level, it is processed layer by layer until the output layer is finally activated. Given adequate concealed units and sufficient data, it has been shown that multilayer perceptron can predict virtually any function to a desired level of accuracy.

From Table 3, it can be inferred that an MPN model was built using two hidden layers. In the model, Y (marketing security of agar growers) was analysed against significant causal variables, namely age (X1), family occupation (X5), annual income (X6), experience in agar cultivation (X7), indebtedness (X9), economic motivation (X10), access to extension services (X14) and risk orientation (X16). The positive value for the output layer, taken from the parameter estimates table for Y, was designated as 0.974 under the hidden layer H (2:3). Within H (2:3), the highest positive interaction among the causal variables was observed as 1.061 under the hidden layer H (1:1), where positive interactions for the variables X1, X6, X10, X14 and X16 were detected under the influence of the same hidden layer, which pretty much refers that these variables generated most number of activated neurons to influence the output variable through the hidden layer H (1:1) and H (2:3). However, consideration of the overall performance of these variables throughout the NN is important to determine masking effects and the direct or indirect influence of other variables on these significant causal variables, which finally determines the contribution of all causal variables to the dependent variable, Y. This is clearly depicted in the normalized graph, where X6 shows the maximum influence on the dependent variable Y. The sequential arrangement of the other X variables is shown as X1 > X10 > X14 > X7 > X16 > X5 > X9, indicating that X6 has the better influence on the dependent variable Y. This establishes a wholesome scenario in which all significant causal variables not only influence the dependent variable but also influence each other in their combined effect on the dependent variable (Table 4 & Fig. 3).

In the polyhedral interactions, exogenous variables are routed through the hidden layers to minimize errors, wherein some input variables were found to generate a significant effect (bold and blue lines) on the output variable. From this, a substantive effect of the input variables in characterizing or scaling up the output (dependent) variable resulted out. As mentioned above, the bold

Table 3. Artificial neural network of predictor variables with respect to marketing security of agarwood growers (Y)

Predictor		Parameter estimates: Predicted												
		Hidden layer 1							Hidden layer 2				Output layer	
		H (1:1)	**H (1:2)**	**H (1:3)**	**H (1:4)**	**H (1:5)**	**H (1:6)**	**H (1:7)**	**H (2:1)**	**H (2:2)**	**H (2:3)**	**H (2:4)**	**H (2:5)**	**Y**
Input layer	(Bias)	**1.026**	.300	.183	.128	-.213	-.673	.083						
	X1	**.225**	-.430	.022	-.222	-.263	.708	-.484						
	X5	-.013	.120	.156	-.050	-.390	.066	-.118						
	X6	**.449**	-.462	.381	-.340	-.126	.448	.602						
	X7	-.142	.127	.122	-.048	-.056	.049	.034						
	X9	-.278	.111	.192	-.167	-.310	-.046	.220						
	X10	**.407**	.202	.192	.474	-.363	-.318	.098						
	X14	**.153**	-.438	.099	.338	.467	-.024	-.210						
	X16	**.093**	.051	.326	.114	.149	.143	-.196						
Hidden layer 1	(Bias)								.447	.100	.787	.461	-.597	
	H (1:1)								.072	.119	**1.061**	.260	-.576	
	H (1:2)								.523	.109	.002	-.051	.303	
	H (1:3)								.006	-.463	.552	-.290	-.322	
	H (1:4)								.499	-.367	.036	-.241	.435	
	H (1:5)								-.096	-.275	-.429	.198	-.324	
	H (1:6)								-.765	.192	-.241	.012	.067	
	H (1:7)								.047	.017	.501	-.356	-.436	
Hidden layer 2	(Bias)													-.306
	H (2:1)													-1.313
	H (2:2)													.175
	H (2:3)													**.974**
	H (2:4)													.007
	H (2:5)													-.697

Table 4. Independent variable importance

Variables	**Importance**	**Normalized importance (%)**
Age (X1)	.291	56.8
Family occupation (X5)	.010	2
Annual income (X6)	.512	100
Experience in agar cultivation (X7)	.019	3.7
Indebtedness (X9)	.009	1.7
Economic motivation (X10)	.120	23.5
Access to extension services (X14)	.024	4.7
Risk orientation (X16)	.014	2.8

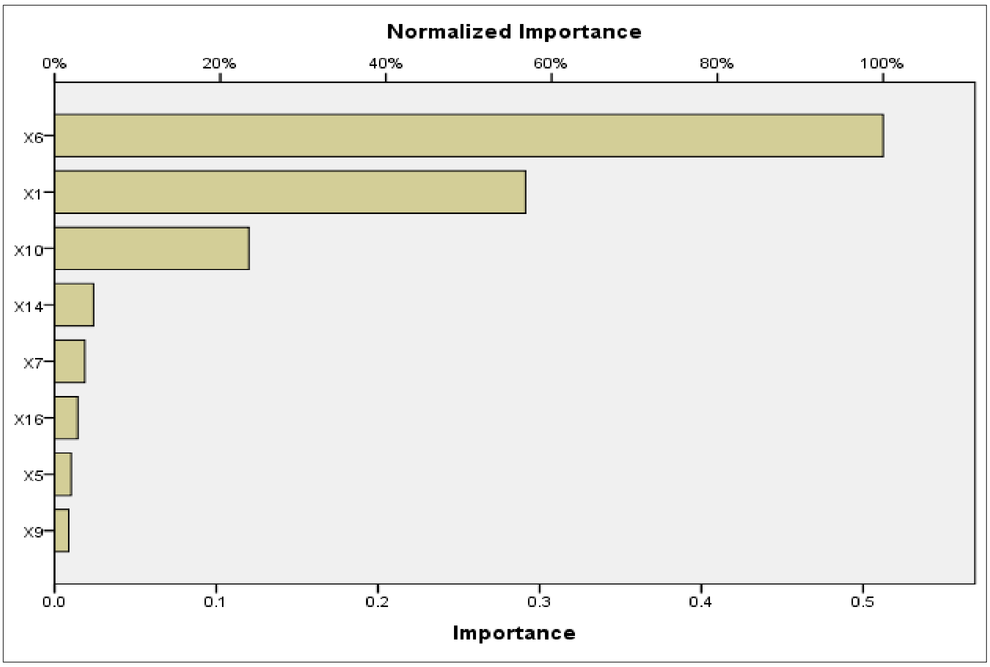

Fig. 3. Normalized importance diagram of predictor variables with respect to marketing security of agar growers (Y).

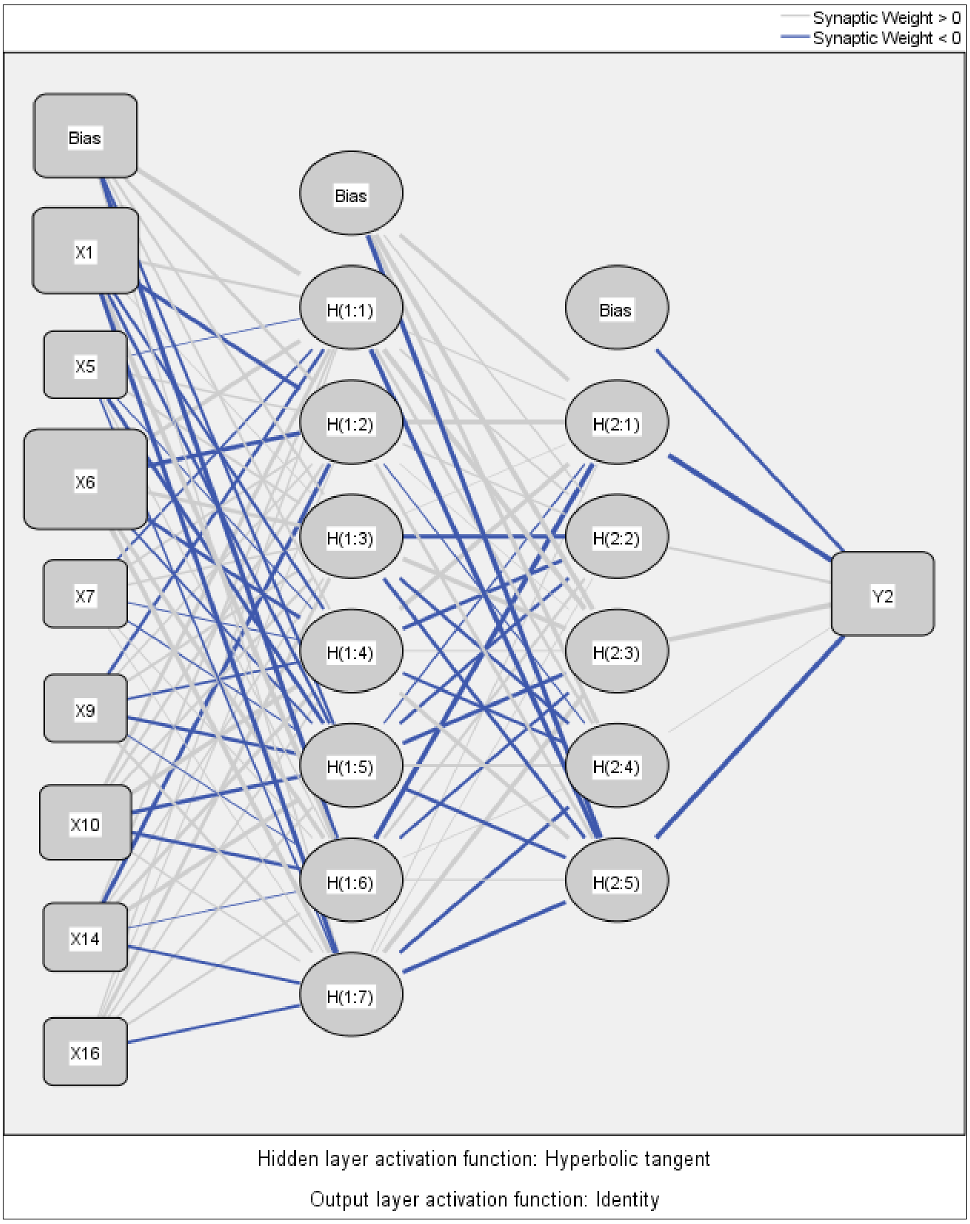

Fig. 4. Artificial neural network of predictor variables with respect to the marketing security of agar growers (Y).

lines depict a stronger influence on Y, the dependent variable (Fig. 4).

Based on Table 4, it was found that the variable annual income (X6) exerted the highest impact on the output variable. To meet daily expenses and cope with high inflation in the present times, agar growers require higher income to maintain a comfortable standard of living with access to essential amenities. Consequently, they explore supplementary means of livelihood by engaging in allied income-generating activities. Due to increase in income from all possible sources, they can wait for obtaining a better market price of agar products. It also indicates that variable X6 contributed an importance value of 0.512 and a normalized importance value of 100 %. Similar findings have been reported in earlier studies (13, 14).

Random forest analysis

Mean decrease accuracy (% Inc MSE) is a measure of model accuracy. If the model accuracy decreases, we need to leave the variable. Increased percentage of mean decrease accuracy (% Inc MSE) can be achieved by assigning values to a variable through random permutation. Higher the value of mean decrease accuracy (% Inc MSE), higher is the variable importance. In this regression analysis model, age (X1) recorded the highest mean decrease accuracy percentage of 30.82, indicating that X1 was the most influential variable affecting marketing security (Y). The influential factors may include freedom in independent decision-making due to accumulated experience, a desire to explore various opportunities, a vast network supported by modern communication facilities and aspirations to earn money through agar cultivation. The results showed that mean square error (MSE) value for the model-marketing security of agar growers (Y) explained by the significant causal variable (X1, X5, X6,X7, X9, X10, X14 and X_{16}) was 0.155. This classifier established that these significant variables could influence the marketing security by 98.48 % (Table 5 & Fig. 5). Similar findings were reported in an earlier study (15).

The most important variables for the ANN (i.e., annual income) and random forest (i.e., age) classifiers differ considerably. This is primarily due to the differences in how these classifiers work. While ranking variables, random forest relies on local, tree-specific splits guided by impurity based variable selection measures, such as information gain, gain ratio and the Gini index. These measures evaluate variables independently at each node

Table 5. Random forest analysis of predictor variables with respect to marketing security of agar growers (Y)

Variable	Mean decrease accuracy (% Inc MSE)	Mean square error (MSE)	Percentage of variable explained
Age (X1)	30.82		
Family occupation (X5)	8.62		
Annual income (X6)	25.23		
Experience in agar cultivation (X7)	6.33	0.155	98.48
Indebtness (X9)	7.47		
Economic motivation (X10)	20.07		
Access to extension services (X14)	4.87		
Risk orientation (X16)	-0.023		

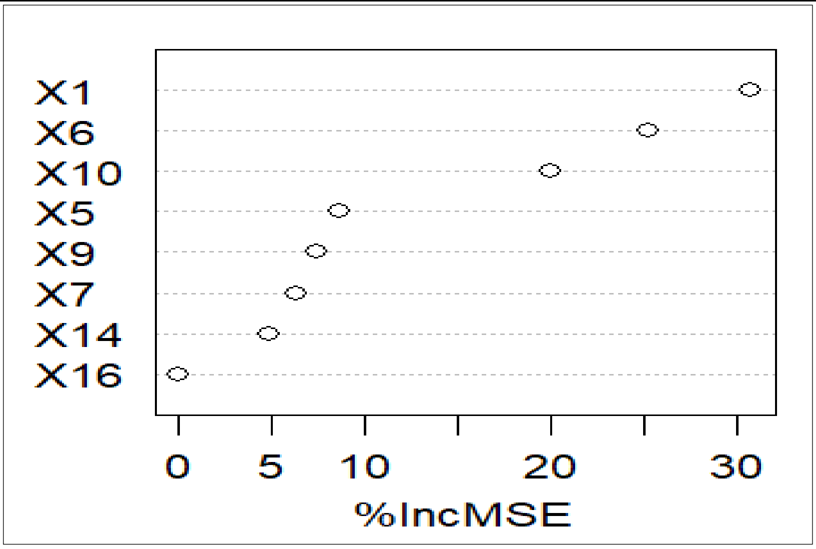

Fig. 5. Random forest analysis of predictor variables with respect to the marketing security of agar growers (Y).

without considering global interactions among them. In contrast, feature ranking in ANN is based on input weights estimated through back propagation, where errors are propagated across layers, enabling the model to capture nonlinear and interdependent feature effects at a global level.

Conclusion

Agar cultivation in Assam offers high economic returns, making it popular among growers. This study, conducted among 420 farmers, revealed that 45.95 % respondents had medium marketing security. Key influencing factors included age, annual income, economic motivation, access to extension services and risk orientation. Among the models, the decision tree showed the highest accuracy (MSE = 0.134), while ANN and random forest identified annual income and age as the dominant predictors. However, validation models used in social science research tend to have some limitations when compared to those in the physical sciences. On a cautionary ground, these well-known limitations may have influenced the present study, although efforts were made to minimize their impact. In addition, the absence of standard universal laws, limited control over variables during analysis, imprecise quantitative measurements and the reflexive human behaviour may also have contributed to model validation constraints. To improve marketing security, growers need better access to extension services, enhanced financial literacy, risk management training and youth-focused support. Structured credit facilities and income diversification strategies can further strengthen market resilience. Although the snowball sampling technique has some limitations regarding its applicability but during the course of research, this sampling technique was used considering the fact that agarwood growers are distantly located and it would have been difficult to identify the agarwood growers by using other sampling techniques. The findings of the study will help policymakers, government agencies, universities, social scientists, NGOs and other research organizations to understand the strengths and weaknesses of agar growers in their efforts and commitment to enhancing marketing security through proper policy advocacy. In addition, the findings will assist policymakers in framing or reframing a solid, scientific agar cultivation and marketing strategy. It was evident that the majority of agar growers possessed a medium level of marketing security. Therefore, government has to create effort in developing well-structured and regulated markets in respective areas to avoid marketing problems and ensure fair prices. Furthermore, government should formulate policies for ease of practicing agar cultivation over larger areas or into other regions of the state. There is enormous potential to link the farmers to the well-structured market through an in-depth analysis of existing marketing channels.

Acknowledgements

The authors are thankful to the agarwood growers of Assam, who provided valuable information, immense assistance and cooperation throughout the entire study.

Authors' contributions

The research topic was conceptualized and refined with the guidance of HKK and the input of RB[1], who played a pivotal role in shaping the overall design and direction of the study. RB[2], UB and ND contributed significantly to the statistical analysis, ensuring methodological rigor and accuracy in data interpretation. JKD, DB and CD were instrumental in the data collection process, overseeing the acquisition and organization of primary data essential to the study's findings. PD provided critical support in structuring and drafting the manuscript, contributing to the framing of arguments and the overall coherence of the research paper. Each author's contribution was integral to the successful completion of this work. [RB [1]-Rituraj Boruah; RB [2]- Rabijita Buragohain]

Compliance with ethical standards

Conflict of interest: The authors declare that there is no conflict of interest.

Ethical issues: None

References

1. Bordoloi S, Ravi N, Modi MK, Jayaraj RS. Genome-wide mining of simple-sequence repeats and development of polymorphic SSR markers in *Aquilaria malaccensis*. Indian J Genet Plant Breed. 2022;82(2):236-39. https://doi.org/10.31742/IJGPB.82.2.15
2. Debnath B, Sil S, Sinha RK, Sinha S. Chromosome number and karyotype of *Aquilaria agallocha* Roxb. (Thymelaeaceae). Cytologia. 1995;60:407-09. https://doi.org/10.1508/cytologia.60.407
3. Dubey S, Dey AN, Roy S, Saha A. Studies on the seed source variation in seed traits of *Aquilaria malaccensis* (Agar). Int J Curr Microbiol Appl Sci. 2020;9(7):1355-69. https://doi.org/10.20546/ijcmas.2020.907.157
4. Kumar R, Das S. Machine learning approaches for agricultural marketing: a review. Agric Inform J. 2021;10(3):45-59.
5. Deloncle A, Berk RD, Andrea F, Ghil AM. Weather regime prediction using statistical learning. J Atmos Sci. 2007;64(5):1619-35. https://doi.org/10.1175/JAS3918.1
6. Giri K, Jayaraj RS, Mishra G, Bordoloi S. *Aquilaria malaccensis* Lamk.: a potential bioresource for shifting cultivation management, livelihood generation and climate change mitigation in Northeast India. Proc Indian Natl Sci Acad. 2020;86(4):1287-89. https://doi.org/10.16943/ptinsa/2020/154984
7. Qiu R, Xu S, Yu Z. Neural networks meet random forests. J R Stat Soc Series B Stat Methodol. 2023;86(5):1435-60. https://doi.org/10.1093/jrsssb/qkae038
8. Government of Assam. Assam Agarwood Promotion Policy, 2020. Gazette notification. Industries and Commerce Department, Assam; 2020.
9. Nair N, Patel M, Gupta S, Singh P. Leveraging neural networks and random forest algorithms for enhanced predictive customer behavior analytics. Int J Artif Intell Adv. 2022;4(2):55-68.
10. Chatterjee R. The estimation of ecological services in terms of energy, climate and knowledge management: the dynamics of social ecology of conservation agriculture. PhD thesis. Mohanpur (West Bengal): Bidhan Chandra Krishi Viswavidyalaya; 2021.
11. Sharma T, Kalita D. Forecasting agricultural commodity prices using machine learning techniques: a case study from Northeast India. J Rural Dev Technol. 2022;15(2):101-15.
12. Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N. WSEAS Trans Circuits Syst. 2009;7(8).
13. Ramadhan AF, Permai SD, Harefa J, Alexander. The comparison of random forest and artificial neural network for customer churn prediction in telecommunication. In: Proceedings of the International Conference on Smart Cities, Automation and Intelligent Computing Systems; Indonesia. 2024. p. 112-18.
14. Draper NR, Smith H. Applied regression analysis. 2nd ed. New York: John Wiley and Sons; 1981.
15. Ahmed M, Borthakur S. Agarwood cultivation and marketing in Assam: challenges and opportunities. Indian J Agrofor. 2020;22 (1):78-85.

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.