ENGINEER - Vol. LVII, No. 03, pp. [35-44], 2024

DOI: https://doi.org/10.4038/engineer.v57i3.7654

Determining the Geographical Authenticity of Sri Lankan Regional Teas with Their Chemical Profile Using Artificial Neural Networks

T.U.S. Peiris and C.K. Walgampaya

Abstract: Characterization of authenticity is a timely requirement in food control and safety. The geographically authenticated food products always realize high market prices and sustainable market share. The present study was conducted to investigate the potential of chemical profile in authenticating geographical differences of Sri Lankan regional black teas using artificial neural networks. Thirteen different geographical origins under eight different *basses* were evaluated and tea samples were chemically analyzed in monthly intervals for a period of one year. Training of feed-forward back propagation multilayer perceptrons resulted in accuracy rates ranging from 46 % - 82 %. Out of eleven chemical parameters considered for the study, Thea Rubigin and Total Colour are the most important discriminators of geographical origins. Other than that, Gereniol, Thea Flavin, Trans – 2 – Hexanol, Total Polyphenol and Amino Acids are also potential discriminators. The study concluded that the differentiation and classification of Geographical Origins (GOs) of Sri Lankan regional black teas is possible using the profile of chemical attributes and applying artificial neural networks. Finally, the data obtained demonstrates that it is possible to assign unknown samples according to their region of origin. Further development of this work could lead to a simple interface that could be used by brokers or exporters to identify the geographical origin of an unknown tea sample.

Keywords: Classification, Geographical Origin, Artificial Neural Network, Sri Lankan Regional Tea

1. Introduction

Determination of food authenticity is one of the most crucial issues in food control and safety, especially the classification of geographical growing origin [1]. The classification of the geographic origin is important for enforcement options for the food industry, protection of the consumer from overpayment and deception.

Tea (*Camellia sinensis* (L.) Kuntze) is one of the most popular beverages in the world. It has been recognized by the medical community as a drink that offers several health benefits. For example, the antioxidant properties of tea have an effect against cancer by inhibiting the formation of cancer-causing compounds [2]. However, chemical composition of tea is concerned, it is a complex product [3] depending on many and diverse factors.

In the Sri Lankan context, Sri Lankan black teas which are renowned world over due to their flavor and quality characteristics [4] are recognized and marketed with the names of the geographical origins where they produced. These geographically authenticated teas realize high market prices and make significant contribution to the economy of the country. However, this authentication is primarily based on sensory evaluation. The problem of subjectivity of sensory evaluation has become a major concern among scientific community [5], [6], [7]. Therefore, proper characterization of authenticity is a timely requirement for Sri Lankan black teas. Thus, characterization and differentiation of tea, based on its geographical origin, is an increasingly active research area.

In addition, attention has also been directed to mapping new geographical locations which have a specific character unique to the region.

On the other hand, one of the most active areas for the application of pattern recognition

Dr. T.U.S. Peiris, B.Sc. (Hons) Agriculture (Peradeniya), M.Phil. (Peradeniya), PhD (Peradeniya), Senior Lecturer in Biostatistics, Biostatistics Unit Wayamba University of Sri Lanka,
Email:ursla@wyb.ac.lk
https://orcid.org/0000-0002-2365-2541
Eng. (Prof.) C.K. Walgampaya, AMIE (Sri Lanka), B.Sc. Eng. (Hons) (Peradeniya), M.Sc.(UofL,KY,USA), Ph.D. (UofL, KY, USA), Professor in Computer Engineering, Department of Engineering Mathematics, Faculty of Engineering, University of Peradeniya.
Email:ckw@pdn.ac.lk
https://orcid.org/0000-0003-2192-474X

techniques is the classification of product brands and quality of origin. Several applications of pattern recognition techniques, mainly principal component analysis (PCA), cluster analysis (CA) and linear discriminant analysis (LDA), were found in literature with respect to establishing the origin of different food products [11]. Upon these techniques, artificial neural networks (ANN) are used more and more frequently in chemistry [8]. ANN classifiers have been successfully implemented for various quality inspection and grading tasks of diverse food products. ANNs are very good pattern classifiers because of their ability to learn patterns that are not linearly separable [8] and concepts dealing with uncertainty, noise and random events [5]. It has been demonstrated in several literature that artificial neural network is a very useful tool for the classification of wine [9], [10], [11].

Several attempts have been made in literature to discriminate tea samples using multivariate statistical techniques with some chemometric procedures [12], [13]. Upon that, [14] have proved that ANN performed better than multivariate classification techniques in differentiating tea types using their metal content.

However, as far as Sri Lankan regional teas are concerned, only one study has conducted to classify geographical origins of Sri Lankan black teas using simple fluorescence-based technique of total luminescence spectroscopy in conjunction with Principal Component Analysis [15]. But this study considered different factories instead of geographical origins. Further, there has been no report found on using either statistical linear classification models or artificial neural networks on ensuring the authenticity of geographical origin of Sri Lankan regional black teas.

Therefore, the objective of this research was to develop a method to confirm the geographical authenticity of Sri Lankan regional teas with some of the chemical attributes using feed-forward Back-propagation multilayer perceptron classifier. Moreover, new geographical origins, which have distinct character specific to the region, were also investigated.

2. Materials & Methods

2.1 Geographical Origins (GOs) of Teas

The same classification of GOs of Sri Lankan black teas, as indicated in Peiris et al. (2017)[16], was used in this study. A brief description of the classification is as follows: Sri Lankan teas have been classified geographically into three major origins as high, medium and low grown teas. Each major origin has been again divided into two categories. High grown teas consist of teas produced in Western & Uva areas. Similarly, medium grown teas consist of Western and Uva medium grown teas. Sabaragamuwa and Ruhuna are the major components that belong to the low grown teas. Western high growns again can be divided into major three types of teas, i.e, Nuwaraeliya, Udapussellawa and Dimbula. Uva high grown teas consist of a major component called Malwatta Valley teas. Accordingly, Nuwaraeliya, Udapussellawa, Dimbula, Western medium grown, Sabaragamuwa and

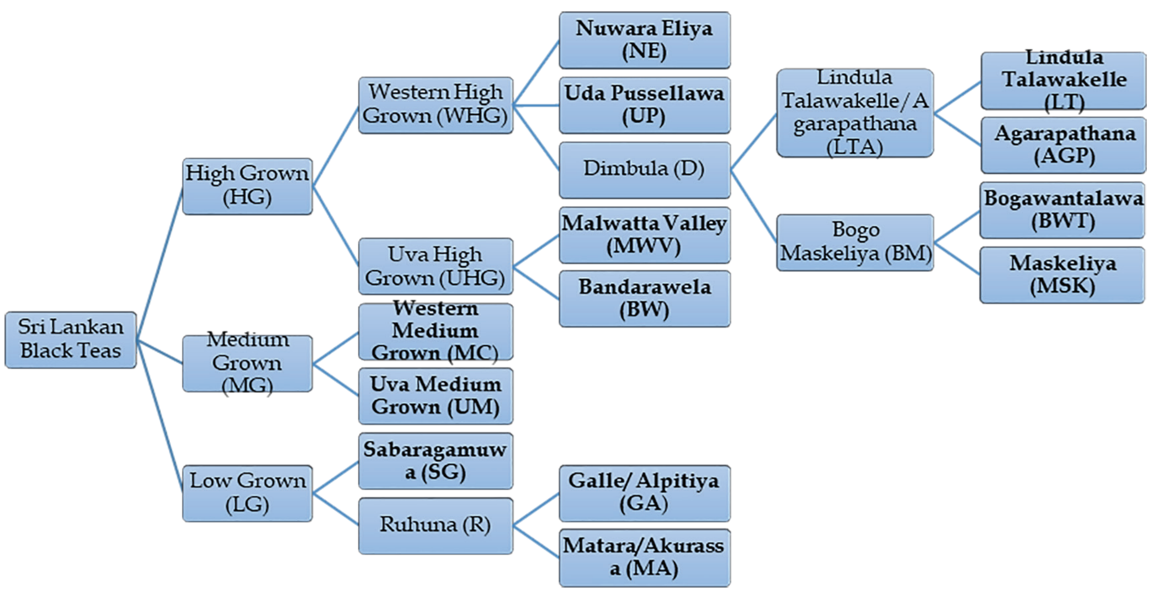

Figure 1 - Relationship of Selected GOs to Broad Classification of Sri Lankan Regional Teas

Ruhuna are the six geographical origins already established in Sri Lanka. However, thirteen GOs having a distinct character unique to the region were considered for the study instead of six established GOs. The selection of additional GOs was based on discussions with tea brokering and exporting companies. The relationship of selected GOs to their broad classification of Sri Lankan regional teas is shown in Figure 1 [16] and selected geographical origins for the study are indicated in bold letters.

2.2 Bases for Classification

Eight different bases were created by amalgamating and omitting individual GOs in order to improve the overall accuracy of GOs. Following presents the GOs included in each base.

Base 1: All thirteen different geographical origins were considered separately.

Base 2: Only five major regions were considered, WHG, UHG, UM, MC and LG.

Base 3: Nine separate regions were considered: LT and AGP were considered as one region (LTA) due to close proximity. BWT and MSK were considered as BM teas. MWV and BW were considered as UHG teas. Teas from MA and GA were considered as R teas.

Base 4: The region UHG was removed from the Base 3 and rest of the eight regions were considered for the classification. The reasons for omitting were due to non-functioning of two factories (Haputale and Pitaratmale) for a considerable number of months within the year in BW region. Therefore, only two data points were available to represent the region. In MWV, the method of manufacturing had changed to orthodox manufacturing from orthodox rotovane manufacturing method.

Base 5: AGP and BW regions were removed from the original regions. MA and GA were amalgamated and considered as R teas.

Base 6: It is same as Base 5 except the omission of UM region.

Base 7: All sub regions under D teas were amalgamated and both MA and GA considered as R teas.

Base 8: It is same as Base 7 except the omission of UHG teas.

2.3 Sampling Strategy

Four representative factories were selected from each geographical origin. Only main tea grades were considered for the chemical analysis. The selection criteria of specific main grades were twofold; the quantity of production in that particular GO and the method of manufacture in the selected factory. Accordingly, Broken Orange Pekoe Fannings (BOPF) or Broken Orange Pekoe (BOP) were collected from the factories adapting orthodox rotovane manufacturing method and Flowery Broken Orange Pekoe (FBOP) or Pekoe were selected from orthodox manufacturing method. The identified grades are key to the quality definition of the region. Most of the factories in the region are producing those grades and the grades cumulatively represent at least 60% of the total production in the region. Tea samples were collected for one year period at monthly intervals.

2.4 Chemical Analysis

Eleven chemical attributes have been applied for classification of regional black teas. Details of the chemical attributes used are given in Table 1.

Simultaneous distillation and extraction (SDE) apparatus and Gas Chromatography Mass Spectrometry (GCMS) were used to estimate concentrations of volatile compounds in regional tea samples. Twenty grams of the tea sample was placed in a single neck round bottom flask and 250 ml of hot water at 60^0C

Table 1 - Attributes Studied for the Chemical Profile of Regional Teas in Sri Lanka

Chemical Compound	Abbreviation	Chemical Compound	Abbreviation
Volatiles (5)		Thea Flavin	TF
Transe 2 Hexanal	T2H	Thea Rubigin	TR
Linalool	L	Amino Acids	AA
Cis-3-Hexanal	C3H	Total Polyphenols	TPP
Methyl Salicilate	MS	Total Color	TC
Gereniol	G	Brightness	B
un-oxidized poly-phenols (Chatechins) (6)			
Chatechin	C	Epi-Gallo Chatechin	EGC
Epi-Chatechin	EC	Epi-Gallo ChatechinGallate	EGCG
Epi-ChatechinGallate	ECG	Caffeine	Cf

*These abbreviations will be used to identify chemical attributes in rest of the text.

was added to the flask. A small flask with 12 ml of Dichloromethane (on a water bath at 40^0C) was connected to the right arm of SDE apparatus. Then 12 ml of Dichloromethane was added along the walls of the U tube of the SDE apparatus. The flask was placed on the heating mantle (100^0C) and linked to the other arm of the SDE apparatus. Then the distillation was started. Then the time that the first bubble of Dichloromethane moves down to the interface line in the U tube was noted. From that point onwards the distillation was continued for half an hour. Then the small flask with condensate of Dichloromethane was disconnected. Using a separation funnel, the Dichloromethane layer was separated. To remove water drops, 2 g of Anhydrous Sodium Sulphate was added and shaken. Then it was kept for half an hour. The solution was filtered through Watman No. 41 filter paper. The solution was concentrated to a small volume (water bath at 40^0C). The dried aliquot was then transferred into a weighted vial and further dried using dry Nitrogen. After that 0.1 µl was injected to the pre-programmed GCMS.

Robert and Smith method [17] was used to analyse Theaflavins, Thearubigins, total Colour and Brightness of regional tea samples. The method explained in ISO 14502 – 1: 2005 [18] was used to analyze total polyphenol content in regional teas.

Amino acids were analyzed in regional tea samples using the method suggested by [19]. Citrate buffer, pH 5 (0.2 M) 0.5 ml was mixed with 1 ml of amino acid solution containing 0.05 to 5.6 µg of amino nitrogen. Then 0.2 ml of the methyl cellosolve ninhydrin solution and 1 ml of the potassium cyanide – methyl cellosolve solution (or 1.2 ml of the potassium cyanide - methyl cellosolve ninhydrin solution) were added. Then the well mixed solution was heated for 15 minutes at 100^0C and cooled for 5 minutes in running tap water. The solution was made up to a convenient volume with ethanol, shaken well and the optical density was determined using a 1 cm glass cell on a Unicam spectrophotometer SP 500 a 570 mµ for all amino acids.

2.5 Multilayer Perceptron

For complex analysis of data and non-linearity between the input and the predicted output, ANN is a powerful tool of execution of various predictive-output tasks [20]. As feed-forward Back-propagation multilayer perceptron (MLP) is a common type of neural network classifier applied to classification problems [21], MLP was used to classify teas into their geographical origins. MLP contains a minimum of three layers: an input layer, output layer and one or more hidden layers.

Here the output of the neuron (Y) after applying the activation function to the linear combination can be explained by,

$$Y = \varphi[w^T x(n) + b] \qquad \ldots(1)$$

where w^T is the transpose of the weight vector w,b is bias vector, $x\,(n)$ is the input vector at instance n,$w^T x\,(n)$ represents the dot product of the weight vector w and the input vector $x\,(n)$,φ is the non-linear activation function and n is the iteration number.

The activation function plays an important role in the training of neural networks. It provides the necessary nonlinearity of the model to be able to learn complex representations [22].

If an activation function is not used in a neural network, then the output signal would simply be a simple linear function which is just a polynomial of degree one. Although a linear equation is simple and easy to solve, its complexity is limited and it does not have the ability to learn and recognize complex mapping from data [23]. There are several activation functions used in MLP such as sigmoid, tanh and rectified linear unit. Here, sigmoid function has been used as activation function.

$$\sigma(x) = \frac{1}{(1+e^{-x})} \qquad \ldots(2)$$

The pattern recognition tools used in this work were as follows in Section 2.5.1.

2.5.1 Back-propagation (BP) Algorithm, Error function (E) and Gradient-Descent Technique (GDT)

The back propagation algorithm is the most popular method for neural network training and it has been used to solve numerous real-life problems [24], [25]. Back propagation is a training procedure for feed–forward neural network that consist an iterative optimization of an error function representing a measure of the performance of the network.

Error function of the Back-propagation algorithm can be defined as [24],

$$E = \frac{1}{2}\sum_{p=1}^{P}\sum_{j=1}^{N_L}\left(t_j - a_j\right)^2 \qquad \ldots(3)$$

where t_j and a_j are the target and actual response values of output neuron j and N_L is the number of output neurons, L being the number of layers, and P representing the number of data points. The back propagation algorithm looks for the minimum of the error function in weight space using the method of gradient descent [25]. The minimization of the error function is carried out using a gradient-descent technique.

$$\Delta w_{ij}(n) = -\eta \frac{\partial E(n)}{\partial w_{ij}(n)} \qquad \ldots(4)$$

[24]

η represents the learning rate and w_{ij} represents the connection weight from unit i in layer l to unit j in layer $l+1$.

Learning rate parameter determines how much the weights can change in response to an observed error on the training set [26]. Though back propagation presents some difficulties in dealing with sufficiently large problems, it was discovered however that the appropriate manipulation of the learning rate parameter during the training process can lead to a very good result [24], [26]. According to Wilson & Martinez (2001)[26], optimum generalization accuracy and training time can be obtained for the range of η (0.1 – 0.001). Therefore, all models were tested for improvement of correctly classified percentages of geographical origins by setting η as 0.01, 0.05, 0.09, 0.1 & 0.3.

2.6 Training of Feed-forward Back Propagation Multilayer Perceptrons

Five different test options were used to validate the model: two cross validation options (10-fold & 15-fold) and three percentage splits (66%, 70% & 75%). In 10- and 15-fold cross validation, 10 and 15 models were developed and the average value was obtained. Three different discretization options (25, 10 & 5 bins) were also checked in addition to feeding continuous data. Discretization is an important feature which can be used to improve the accuracy of learning algorithms. Attribute selection to build the neural network model was also done using filter and wrapper methods apart from the all chemical attributes. Two methods of discretization (equal length and equal frequency) were tested at each level of bins. A mix of continuous and discretized chemical attributes were also tested in two different ways. Five different learning rates (0.001, 0.005, 0.009, 0.1 & 0.3) were also tested. Therefore, 240 different models were tested for each base to find out the best model.

Attributes consisting of a high discriminatory power were selected by evaluating commonly used two filtering methods called filter (Info Gain Attribute Eval) and wrapper (Wrapper Subset Eval). Info Gain Attribute Eval evaluates the worth of an attribute by measuring the information gain with respect to the class. It uses an attribute evaluator and a ranker to rank all the features in the dataset. Wrapper Subset Eval evaluates attribute sets by using a learning scheme. Wrapper method uses subset evaluator to create all possible subsets from the feature vector. Then it uses a classification algorithm to induce classifiers from the features in each subset. Then it will consider the subset of features with which the classification algorithm performs the best. Cross validation is used to estimate the accuracy of the learning scheme for a set of attributes [27]. Filtering method which would give the highest accuracy rate was selected and attributes which lie in first five ranks were used as inputs to the neural network. All input variables were normalized before feeding to the neural network. Data quality was improved by executing all necessary data pre-processing techniques such as filtering, eliminating missing values, etc.

2.7 Waikato Environment for Knowledge Analysis

Waikato Environment for Knowledge Analysis (WEKA), an open source software platform that provides an integrated environment for machine learning, was used for the analysis.

2.8 Performance Measures to Evaluate the Classifier

To evaluate the performance of the model which was built with multilayer perceptron algorithm, four performance measures, i.e., accuracy, precision, recall and f-measure were used. All performance measures are calculated based on following four parameters [28].

<table>
<tr><td></td><td colspan="3">Predicted Class</td></tr>
<tr><td rowspan="3">Actual Class</td><td></td><td>Class = yes</td><td>Class = no</td></tr>
<tr><td>Class = yes</td><td>True Positive (TP)</td><td>False Negative (FN)</td></tr>
<tr><td>Class = no</td><td>False Positive (FP)</td><td>True Negative (TN)</td></tr>
</table>

Accuracy is the ratio of correctly predicted observations to the total number of observations.

$$Accuracy = \left(\frac{TP+TN}{TP+FP+TN+FN}\right) * 100 \qquad \ldots(5)$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \left(\frac{TP}{TP+FP}\right) * 100 \qquad \ldots(6)$$

Recall or sensitivity is the ratio of correctly predicted positive observations to all observations in actual class

$$Sensitivity = \left(\frac{TP}{TP+FN}\right) * 100 \qquad \ldots(7)$$

F1 is the weighted average of precision and recall.

$$F1 = \left(\frac{2\,P*R}{P+R}\right) \qquad \ldots(8)$$

3. Results

Neural network models that resulted in the highest accuracy rates were obtained for 10 or 15–fold cross validation options rather than percentage split options for all eight bases (Table 2). The learning rate of 0.1 depicted the highest accuracy rates for all bases. It lies within the optimum generalization accuracy range as cited in [26]. In addition, the momentum of 0.2 gave the highest accuracy for all bases. Further to that, multilayer perceptrons with a single hidden layer shows the highest accuracy for all bases.

According to Table 2, Wrapper method of attribute selection most frequently (5 out of 8 bases) yields the highest accuracy rates than the filter method. Filter method (InfoGainAttributeEval) of attribute selection shows the highest accuracy rate for Bases 1, 2 and 5 while wrapper method (WrapperSubsetEval) of attribute selection shows the highest accuracy rate for the rest 5 bases. As more than 5th rank of attributes does not show significant contribution to improve accuracy only attributes with first five ranks were used for training neural network.

Results showed that TR and TC are the most potential attributes contributed to the discrimination of geographical origins. It has been established that the coloured oxidation products extracted in the liquor called thearubigins and theaflavins are responsible for almost all colour variation in the liquor [29], [30]. Apart from that G, TF, T2H, TPP and AA have also contributed to discriminate GOs.

Table 2 depicts that out of 8 different bases, Base 2 is the best fitted model with 82% accuracy rate. However, it has only 5 GOs and include only two individual GOs (UM & MC). The second highest accuracy rate (75%) was observed in Base 8 with 7 GOs. Out of that 5 are individual GOs. Accuracy rates of Base 3 – Base 8 have ranged from 58% – 67%. The model fitted for 13 individual GOs (Base 1) has resulted in the lowest accuracy rate of 46% compared with the other bases.

Performance measures in eight different bases are given in Table 3. Except Base 1, all the other bases f – measure is >50%, ranging from 0.552 to 0.793. The highest f – measure is shown in Base 2 (0.793) followed by Base 8 (0.71). Out of eight bases, 80% of positive instances in Base 2 are true to type while in Base 8 it is 70%. Therefore, Base 2 and Base 8 can be considered as satisfactorily fitted neural networks.

Next, the broad classification of regional teas is concerned (Base 2), except UvaH, f-measure for all the other GOs were found to be satisfactory (0.68 – 0.91). Reason for poor f-measure in UvaH is due to very low specificity (0.23). Out of the total instances that classified as HG &

Table 2 - Parameters and the Accuracy Rate of the Best Fitted Model for Eight Different Bases of Geographical Origins

Base	GOs	Validation	Most important attributes	RMSE	Nodes	Accuracy Rate (%)
1	13	10F	TR,TC, TF,G,TRTF	0.229	9	46
2	5	10 F	TR,TC, G, T2H,TRTF	0.245	5	82
3	9	15 F	TR,TC,G,AA,TRTF	0.253	7	58
4	8	10 F	TR,TC,G,TPP,AA	0.247	6	64
5	10	10 F	TR,TC, G, TF,TRTF	0.240	7	59
6	9	10 F	TR,TC,G,TPP,AA	0.251	7	59
7	9	15 F	TR,TC,G,T2H,TF	0.234	7	67
8	7	15 F	TR,TC,G,T2H,MS	0.231	6	75

*LR & M indicate learning rate and momentum respectively

LG, around 85% was found to be true to the type. Furthermore, 95% of teas originally in HG & LG categories, have been classified true to the type.

Amalgamating MA & GA GOs as R (Base 3), f-measure has improved to a satisfactory level (0.76). In addition, 91% of R teas, have been classified as true to the type. Similarly, out of the total instances that classified as R teas, 66% was found to be true to type. Moreover, amalgamation of BWT & MSK as BM teas has also improved f-measure to a significant level (0.62). However, amalgamation of LT & AGP, and MWV & BW has not contributed to a satisfactory level of f-measure.

Compared to the Base 3, removal of UvaH teas from the model (Base 4) has caused to reduce the f-measure of SG teas to zero. In contrast, it has increased the f-measure of BM (0.71) and LTA (0.57). Removal of AGP and BW, and considering BM in its individual GOs (Base 5) compared to Base 3, has increased the f-measure in UP and MWV by 0.26 and 0.16, respectively. According to Base 6, removal of UM has caused to increase of f-measure in MC up to 0.8 compared to the 0.67 in Base 5. In Base 7, amalgamation of all GOs (LT, AGP, BWT and MSK) belongs to D improves f-measure to 0.79.

Table 3 - Performance Measures in Eight Different Bases

Base	TP Rate	FP Rate	Precision	Recall	F–Measure
1	0.46	0.045	0.431	0.46	0.443
2	0.818	0.093	0.797	0.818	0.793
3	0.58	0.065	0.556	0.58	0.552
4	0.644	0.065	0.582	0.644	0.604
5	0.587	0.053	0.569	0.587	0.568
6	0.594	0.059	0.575	0.594	0.569
7	0.668	0.078	0.603	0.668	0.619
8	0.75	0.065	0.705	0.75	0.71

Table 4 - Performance Measures of Individual Geographical Origins of the Best Fitted Model for Eight different Bases

Base	PM	NE	UP	LT	AGP	BWT	MSK	MWV	BW	UM	MC	SG	MA	GA
1	P	0.77	0.33	0.3	0.36	0.38	0.33	0.36	0	0.72	0.58	0.46	0.51	0.44
	R	0.83	0.27	0.25	0.35	0.5	0.43	0.38	0	0.79	0.65	0.38	0.58	0.5
	F	0.8	0.3	0.27	0.35	0.43	0.37	0.37	–	0.75	0.61	0.41	0.54	0.47
2	P	0.84						0.61		0.77	0.7	0.86		
	R	0.95						0.23		0.77	0.67	0.96		
	F	0.89						0.33		0.77	0.68	0.91		
3	P	0.89	0.31	0.45		0.55		0.47		0.77	0.6	0.37	0.66	
	R	0.83	0.19	0.56		0.72		0.24		0.77	0.67	0.15	0.91	
	F	0.86	0.23	0.5		0.62		0.31		0.77	0.63	0.21	0.76	
4	P	0.91	0.41	0.59		0.65		–	–	0.71	0.61	0	0.63	
	R	0.85	0.27	0.54		0.79		–	–	0.73	0.65	0	0.96	
	F	0.88	0.33	0.57		0.71		–	–	0.72	0.63	–	0.76	
5	P	0.9	0.49	0.36	–	0.41	0.46	0.51	–	0.76	0.67	0.42	0.64	
	R	0.92	0.5	0.27	–	0.5	0.4	0.43	–	0.79	0.67	0.17	0.9	
	F	0.91	0.49	0.31	–	0.45	0.43	0.47	–	0.78	0.67	0.24	0.74	
6	P	0.79	0.49	0.45	–	0.43	0.43	0.63	–	–	0.75	0.44	0.67	
	R	0.85	0.42	0.4	–	0.54	0.43	0.4	–	–	0.85	0.15	0.94	
	F	0.82	0.45	0.42	–	0.48	0.43	0.49	–	–	0.8	0.22	0.78	
7	P	0.85	0.41	0.71				0.47	0	0.74	0.65	0.5	0.66	
	R	0.92	0.27	0.89				0.38	0	0.83	0.58	0.15	0.95	
	F	0.88	0.33	0.79				0.42	-	0.78	0.62	0.23	0.78	
8	P	0.89	0.62	0.84				–	–	0.79	0.62	0.17	0.65	
	R	0.88	0.27	0.94				–	–	0.71	0.77	0.02	0.94	
	F	0.88	0.38	0.89				–	–	0.75	0.69	0.04	0.77	

PM indicates performance measures, P,R& F indicate precision, recall &F-measure respectively

In Base 8, removal of UvaH teas compared with Base 7, f-measure has increased to 0.89. In addition, results showed that 95% of teas belonging to D have classified true to the type, and out of total instances classified as D teas, 84% is true type.

4. Discussion

In this study, neural network models with one hidden layer gives the highest accuracy rate in all bases. This confirms the results of the study done to recognize the country of origin of tea using metal contents [14].

Due to high sensitivity values (83% – 92%) across all bases, NE shows its uniqueness as a GO. In addition to NE, UM (71% – 83%) and MC (65% – 85%) teas can also be accurately discriminated out of the individual GOs. Interestingly, UM has not previously established as a region having a unique character. According to the results of this study, UM is introduced as a new regional tea type having specific character.

Though UP is a well-established regional tea type in Sri Lanka producing exquisitely tangy teas, it is shown that it cannot be accurately identified using chemical attributes. One reason may be of changing the method of manufacturing from orthodox rotovane to pure orthodox in some tea factories in the region. It was noticed that in most of the bases, UP teas have been classified as D teas.

As an already established tea type producing refreshingly mellow teas, this study confirmed the uniqueness of D due to high sensitivity values (89% – 94%). When individual GOs of D are concerned, LT & AGP cannot be discriminated accurately. However, BWT & MSK show a marginal potential. These two GOs, together which called as BM teas, depict high discriminating potential due to high sensitivity values (72% – 79%).

Though UvaH teas have been classified as a unique GO under regional tea classification by Sri Lanka Tea Board, it shows very low discriminating potential. When individual GOs are compared, all BW teas have been classified into other regions with a 60% to D teas. 62% of MWV teas have been dispersed to all other GOs without concentrating to one region. One reason for not showing is its uniqueness that all factories of MWV region have been converted to pure orthodox method of manufacturing rather than its conventional orthodox rotovane method of manufacturing during the period of data collection except flavour season.

LG teas show very high discriminability with 96% of sensitivity however, SG shows very low discriminating potential across all bases. More than 50% of SG teas have been classified into R teas. Though, MA & GA show marginal potential as individual GOs, amalgamation of two regions as R teas depict higher discriminability (90% – 96% of sensitivity).

Finally results reveal that NE, UM, MC, D, BM and R are the GOs which show high discriminating potential.

Results are also evident that TR and TC are the most important discriminators under any base concerned.

5. Conclusions

In conclusion, out of eleven chemical parameters considered for the study, TR and TC are the most important discriminators of GOs. Other than that G, TF, T2H, TPP and AA are also worth to analyze for discrimination. This study clearly shows that the differentiation and classification of GOs of Sri Lankan regional black teas is possible using the profile of chemical attributes and applying artificial neural networks. As individual GOs, differentiation of NE, UM and MC are greatly satisfactory. UM is a new GO identified by this study which has a specific character unique to the region. D, BM and R GOs are also satisfactorily classified using chemical attributes. Though it is a well-established type of regional teas in Sri Lanka, UP has not shown a potential of authenticity as a region. The study also concludes that D teas have a very high power of authentication as a region of origin. However, individual GOs of D have poor geographically unique identity. But, BM teas show a significant specialty to the region. A large number of teas from MWV & BW regions would be needed to establish a satisfactory classification as low number of samples were obtained during data collection. The results of MWV do however give an insight into uniqueness as a GO. Results suggest that LG teas have a high power of authenticity while individual LG regions (SG, MA & GA) have a poor geographical identification. However, R teas have a high potential of using as an authenticated tea. Finally, the data obtained demonstrates that it is possible to assign unknown samples according to their region of origin. Further development of this work could lead to a simple interface that could be used by brokers or exporters to identify the geographical origin of an unknown tea sample.

Acknowledgements

Authors would like to greatly acknowledge the Tea Research Institute of Sri Lanka for financial support in conducting the research. Gratitude is also extended to all Estate Managers for providing tea samples and other required information for the study. A special thanks goes to Mr. H K K Deshapriya, former Technical Officer, Biometry Unit, Tea Research Institute of Sri Lanka for his assistance in collection and analysis of regional tea samples. Thanks are also expressed to the Head and the staff of the Biochemistry Division for granting permission to use equipment and the support provided during the sample analysis.

References

1. Jarroop, Z., Mencer, A., and Tuah, L., "Classification of Geographical Growing Origin of Food Commodities,"*17th IPC Quality Community, Lombok, Indonesia.*

2. Moreda-Piñeiro, A., Fisher, A., and Hill, "The Classification of Tea According to Region of Origin using Pattern Recognition Techniques and Trace Metal Data," *Journal of Food Composition and Analysis*, Vol. 16, No. 2, pp. 195–211, 2003, doi: 10.1016/S0889-1575(02)00163-1.

3. Cimpoiu, C., Cristea, V. M., Hosu, A., Sandru, M., and Seserman, L., "Antioxidant Activity Prediction and Classification of some Teas using Artificial Neural Networks," *Food Chem*, Vol. 127, No. 3, pp. 1323–1328, Aug. 2011, doi: 10.1016/j.foodchem.2011.01.091.

4. Yamanishi, T., Botheju, W. S., and M. Jayanthi De Silva, "An Index for Assessing the Quality of Uva Seasonal Black Tea", *Sri Lanka J. Tea sci*, Vol. 58, No. 3, pp. 40-49, 1989.

5. Debska, B., and Guzowska-Świder, B., "Application of Artificial Neural Network in Food Classification," *Anal Chim Acta*, Vol. 705, No. 1–2, pp. 283–291, Oct. 2011, doi: 10.1016/j.aca.2011.06.033.

6. Gawel, R., and Godden, P. W., "Evaluation of the Consistency of Wine Quality Assessments from Expert Wine Tasters," *Aust J Grape Wine Res*, Vol. 14, No. 1, pp. 1–8, Apr. 2008, doi: 10.1111/j.1755-0238.2008.00001.x.

7. Rossi, F., "Assessing Sensory Panelist Performance using Repeatability and Reproducibility Measures," *Food Quality and Preference,* Vol. 12, pp 467-479, 2001.

8. Jwo, D. J., Biswal, A., and Mir, I. A., "Artificial Neural Networks for Navigation Systems: A Review of Recent Research," *Applied Sciences (Switzerland)*, Vol. 13, No. 7. MDPI, Apr. 01, 2023. doi: 10.3390/app13074475.

9. Kruzlicova, D., Mocak, J., Balla, B., Petka, J., Farkova, M., and Havel, J., "Classification of Slovak White Wines using Artificial Neural Networks and Discriminant Techniques," *Food Chem*, Vol. 112, No. 4, pp. 1046–1052, Feb. 2009, doi: 10.1016/j.foodchem.2008.06.047.

10. Katumullage, D., Yang, C., Barth, J., and Cao, J. "Using Neural Network Models for Wine Review Classification," *Journal of Wine Economics*, Vol. 17, No. 1, pp. 27–41, Feb. 2022, doi: 10.1017/jwe.2022.2.

11. Jana, D., Bhunia, P., Das Adhikary, S., and Mishra, A., "Analyzing of Salient Features and Classification of Wine Type Based on Quality through Various Neural Network and Support Vector Machine Classifiers," *Results in Control and Optimization*, Vol. 11, p. 100219, Jun. 2023, doi: 10.1016/j.rico.2023.100219.

12. Valera, P., Pablos, F., and GustavoGonzhlez, A., "Classification of Tea Samples by their Chemical Composition using Discriminant Analysis," *Talanta*, Vol.43, No.3, pp 415-419, 1996.

13. Liu, X., Van Espen, P., Adams, F., Yan, S. H., and Vanbelle, M., "Classification of Chinese Tea Samples According to Origin and Quality by Principal Component Techniques," *Analyticachimicaacta,* Vol.200, pp. 421-430, 1987.

14. Fernández-Cáceres, P. L., Martín, M. J., Pablos, F., and González, A. G., "Differentiation of Tea (Camellia sinensis) Varieties and their Geographical Origin According to their Metal Content," *J Agric Food Chem*, Vol. 49, No. 10, pp. 4775–4779, 2001, doi: 10.1021/jf0106143.

15. Seetohul, L. N., Scott, S. M., O'Hare, W. T., Ali, Z., and Islam, M., "Discrimination of Sri Lankan Black Teas using Fluorescence Spectroscopy and Linear Discriminant Analysis," *J Sci Food Agric*, Vol. 93, No. 9, pp. 2308–2314, Jul. 2013, doi: 10.1002/jsfa.6044.

16. Peiris, T. U. S., Walgampaya, C. K., Thattil, R. O., and Abeysinghe, I. S. B., "Classification of Geographical Origin of Sri Lankan Back Tea using Discriminant Function Analysis," *Tropical Agricultural Research,* Vol. 29, No.1 pp. 12 – 24 2017.

17. Roberts, E. A. H., and Smith, R. F., "The Phenolic Substances of Manufactured Tea", *JSci Food Agric*, Vol. 14, No. 10, pp. 689-700, 1963.

18. ISO 14502-1,"Determination of Substances Characteristic of Green and Black Tea," *ISO*, 2005.

19. Yemm, E. W., Ocking, E. C., and Rickets, R.E., "The Determination of Amino-acids with Ninhydrin," *Analyst*, pp.209-214, 80(948),1955.

20. Desai, M., and Shah, M., "An Anatomization on Breast Cancer Detection and Diagnosis Employing Multi-Layer Perceptron Neural Network (MLP) and Convolutional Neural Network (CNN)," *Clinical eHealth*, Vol. 4. KeAi Communications Co., pp. 1–11, Jan. 01, 2021. doi: 10.1016/j.ceh.2020.11.002.

21. Irmak, B., and Gülcü, Ş., "MANAS Journal of Engineering Training of the Feed-Forward Artificial Neural Networks using Butterfly Optimization Algorithm," Vol. 9, No. 2, pp. 160–168, 2021, doi: 10.51354/mjen.

22. Rasamoelina, A.D., Adjailia, F., and Sinčák, P., "A Review of Activation Function for Artificial Neural Network," *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp.281-286, 2020.

23. Sharma, S., Sharma, S., and Athaiya, A. "Activation functions in Neural Networks" *Towards Data Sci,* Vol. 6, No. 12, pp. 310-316. 2020.

24. Moreira, M., and Fiesler, E., "Neural Networks with Adaptive Learning Rate and Momentum Terms," 1995. Available: https://publications.idiap.ch/

25. R. Rojas ; R. Rojas and ; R Rojas, "A Systematic Introduction," *Neural Networks*, *622*, 61068-4, 1996.

26. Wilson, D. R., and Martinez, T. R., "The Need for Small Learning Rates on Large Problems," in *Proceedings of the International Joint Conference on Neural Networks*, 2001, pp. 115–119. doi: 10.1109/ijcnn.2001.939002.

27. "Attribute Evaluators," 2008. [Online]. Available: https://dataminingntua.files.wordpress.com/2008/04/weka-select-attributes.pdf

28. Sokolova, M., Lapalme, and G., "A Systematic Analysis of Performance Measures for Classification Tasks," *Inf Process Manag*, Vol. 45, No. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

29. Wickremasinghe, R. L., "Studies on the Quality and Flavour of TeaI-Thepolyphenols and Low Volatile Compounds," *Tea Quarterly*, Vol. 36, pp. 59-63, 1965.

30. Roberts, E. A. H., "The Estimation of Theaflavins and Thearubigins in Made Teas," *Tea Quarterly*, Vol. 29, pp. 99–10, 1958.