

# Classical Image Based Classification of Coffee Beans on Their Botanical Origins in Tongo and Wambara, Benishangul Gumuz, Ethiopia

Lamessa Dingeta Olike<sup>1</sup>, Dejenie Demisse Amberber<sup>2</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Assosa University, Benishangul Gumuz, Ethiopia

<sup>2</sup>Department of Physics, Assosa University, Benishangul Gumuz, Ethiopia

## Email address:

[lms.dngt@gmail.com](mailto:lms.dngt@gmail.com) (L. D. Olike), [dejeniedemisse@yahoo.com](mailto:dejeniedemisse@yahoo.com) (D. D. Amberber)

## To cite this article:

Lamessa Dingeta Olike, Dejenie Demisse Amberber. Classical Image Based Classification of Coffee Beans on Their Botanical Origins in Tongo and Wambara, Benishangul Gumuz, Ethiopia. *Science Journal of Circuits, Systems and Signal Processing*. Vol. 9, No. 2, 2020, pp. 42-48. doi: 10.11648/j.cssp.20200902.12

**Received:** June 19, 2020; **Accepted:** July 29, 2020; **Published:** August 13, 2020

---

**Abstract:** Ethiopia is a homeland of coffee. Coffee is a major export commodity of Ethiopia, which has a significant role in earning foreign currency. This research was conducted with the objective of developing an appropriate computer routine algorithm that can characterize different varieties of Beneshanguel coffee based on their growing region. Imaging techniques were employed to automatically classify the coffee bean samples according to provenance in Beneshanguel (Tongo and Wombara) which corresponds to their botanical origins. Important coffee bean features, namely, color, shape and size and texture were extracted from 100 images (50 images from each location). For the purpose of classification, altogether 24 features (12 colors, 6 shapes and size and 6 textures) were extracted from images of the coffee samples from the two locations. Artificial neural network (ANN) was employed to automatically categorize the coffee beans according to their provenance. We have compared classification approaches of Neural Network classifiers were employed based on the features used for color, morphology (shapes and size), texture, and the combination of morphology and color respectively. To evaluate the classification accuracy, from the total of 100 sample images of the training 70% (70 images), validation 20% (20 images) and testing 10% (10 images) data. Classification scores of 93%, and 99.3% were achieved for color, morphology, texture and a combination of morphology and color features, respectively. The classification results of the network indicated that morphology and a combination of morphological and color features exhibited the highest accuracy. In conclusion, the results of this study have revealed that imaging technique could be used as the most effective method to determine coffee bean qualities for export. However, it is suggested that the repeatability of this coffee quality testing method be validated using a large data set before employing the algorithm for the purpose of classifying coffee beans as a daily routine.

**Keywords:** Beneshangul Coffe, Coffee Beans, Classification, Image Analysis, Neural Networks

---

## 1. Introduction

### 1.1. Background of the Study

Coffee is an edible commodity. It is widely used as a beverage but now a day's its use as input in some food processing industries is increasing [3]. For instance, it is used as a flavoring to various pastries, ice-creams, chocolate, etc.

There are different types of coffee in the world. Among different types of coffee, the major economic species are coffee Arabica and coffee Robusta. Arabica accounts 80% of the world

coffee trade, and Robusta most of the remaining 20%. Coffee Liberica and Excelsa together supply less than 1% [19].

The origins of the coffee crop can be traced back to the Ethiopian highlands for coffee Arabica and the forest of West and Central Africa for coffee Robusta (Canephora). Coffee was well established as a beverage in Yemen by the 14<sup>th</sup> other Middle Eastern countries in the 15<sup>th</sup> century, from where it spread to century and across Arabian Sea to India. Today coffee is widely cultivated and used throughout the tropics [12].

Ethiopia has a suitable environment to grow all Arabica coffee varieties. Currently, only Coffee Arabica is grown in

Ethiopia. Other coffee species are not cultivated yet. Ethiopia being the home of Arabica coffee, the first coffee was discovered from south-western massive highlands of Ethiopia called Kaffa, more specifically from a district called Buno. In Ethiopia, coffee production is concentrated in the Oromia and Southern regions of the country, though the majority of Ethiopian regions are still suitable for coffee growth [12]. Ethiopia is not only the icon of coffee, but it thrives on coffee and people drink coffee regularly in every part of the country. Coffee is closely associated with the Ethiopian culture. Most people in the country start their day by taking a cup or two of coffee in the morning. Coffee ceremony, the tradition of serving coffee in Ethiopia is unique.

### 1.2. Statement of the Problem

In agricultural industry, quality assessment and varieties of agricultural products create the main problems. Nowadays, the quality and varieties of grain seed have been determined manually through a visual inspection by experienced technicians. So it requires high degree of accuracy to satisfy customer need of high level of quality, as well as correctness for a non-destructive quality evaluation method, which is proposed, based on image processing [13]. In different parts of Ethiopia, different coffee varieties are grown. The coffee beans produced in different parts of the country have been prepared for local consumption and for international market by traditional inspection, which makes it subjective and non effective. Identifications are based on physical properties like color, size, shape and flavor all of which are frequently examined by human inspection.

The main interest of this research is to develop computer algorithm that can classify Benishangul region coffee of two botanical origins. Coffee beans that are produced in different parts of the country have distinct physical properties. One of the best ways to classify coffee bean based on their botanical origin is image analysis. Image analysis uses to prepare appropriate model that uses for multipurpose techniques like sorting, classification of their botanical regions and variety classification.

Once the algorithm is developed, it is possible to do classification of variety, identification of growing region and sorting of coffee beans. Using traditional method, one cannot distinguish all the features that computer vision can detect. Due to this reason, automate classification, identification and sorting of coffee bean will not be subjective and non efficient. On the other hand, the use of technology based model will make the country more competent and the coffee to be accepted without doubt, all over the world.

Therefore, this thesis work will initiate a model for Ethiopian coffee variety classification which is consistent, efficient and cost effective by exploring the technology of image analysis.

In agricultural industry, quality assessment and varieties of agricultural products create the main problem. Nowadays the qualities and varieties of grain seed have been determined manually through a visual inspection by an experienced technician. However, it requires high degree of accuracy to satisfy customers' need of high level of quality, as well as correctness for a non-destructive quality evaluation method,

image processing is a preferred method [13].

Ethiopian coffee is an important source of coffee genetic resources for the world coffee industry. As a matter of fact, Ethiopia is the only center of origin and diversity of Arabica coffee (*Coffea arabica* L) [15].

In Ethiopia, coffee grows over a wide range of agro-ecology zones and geographical regions. [7] put it, size and shape difference of coffee beans were influenced by botanical variety and environmental growth circumstances. Coffee grows under diverse environmental condition ranging from 550 to 2600 m above sea level, with annual rainfall from 1000 to 2000mm, and minimum and maximum temperature ranges of 8 to 15°C, and 24 to 31°C, respectively. Coffee requires deep, well drained, loamy and slightly acidic soils [7].

Ethiopian coffee is designated by geographical place name, letters (to tell sub regions) and grade 1 through 9 to tell quality of the coffee and also all coffee divide into four large groups as, commercial washed, commercial unwashed, specialty washed and specialty unwashed [9].

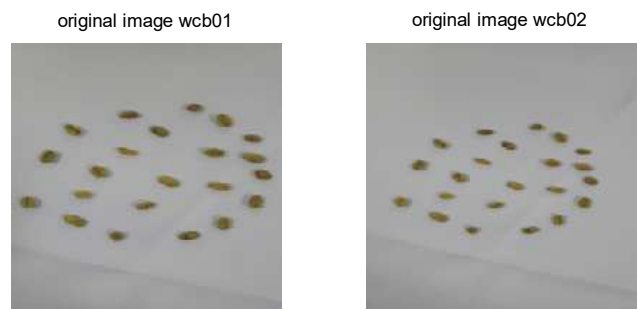
## 2. Results and Discussion

In this study, MATLAB version R2013a has been used to implement all image processing and analysis algorithms. All steps of MATLAB scripts that have been used to perform the above pre- processing steps are clearly illustrated in Appendix.

Results on segmentation, feature extraction and classification are given followed by feature extraction and finally, Benishangule coffee bean classification are made based on ANN in order to permit comparison of our system with those introduced in the literature.

### *Pre-Processing Images of Coffee Beans*

Samples will be collected from selected coffee growing parts of the Benishangul region (Tongo and Wombera). From the local market five Quintals (one Quintal contains 100kg) will be randomly selected, from each selected places and 100gm will be taken from each Quintal (1kg from each of selected part of the country). From the collected samples of 100gm, 50gram will be used for image analysis and the rest are used as reference. 10 images will be taken per each Quintal, which means 50 image from each selected coffee growing part of the country and totally 100 images will be captured and out of this 60% to 70% of the image will be used for practice purpose and the rest will be used for thesis work. The coffee bean images shows in figure 1 are the first experiment with a matrix size of 256×256 pixels.



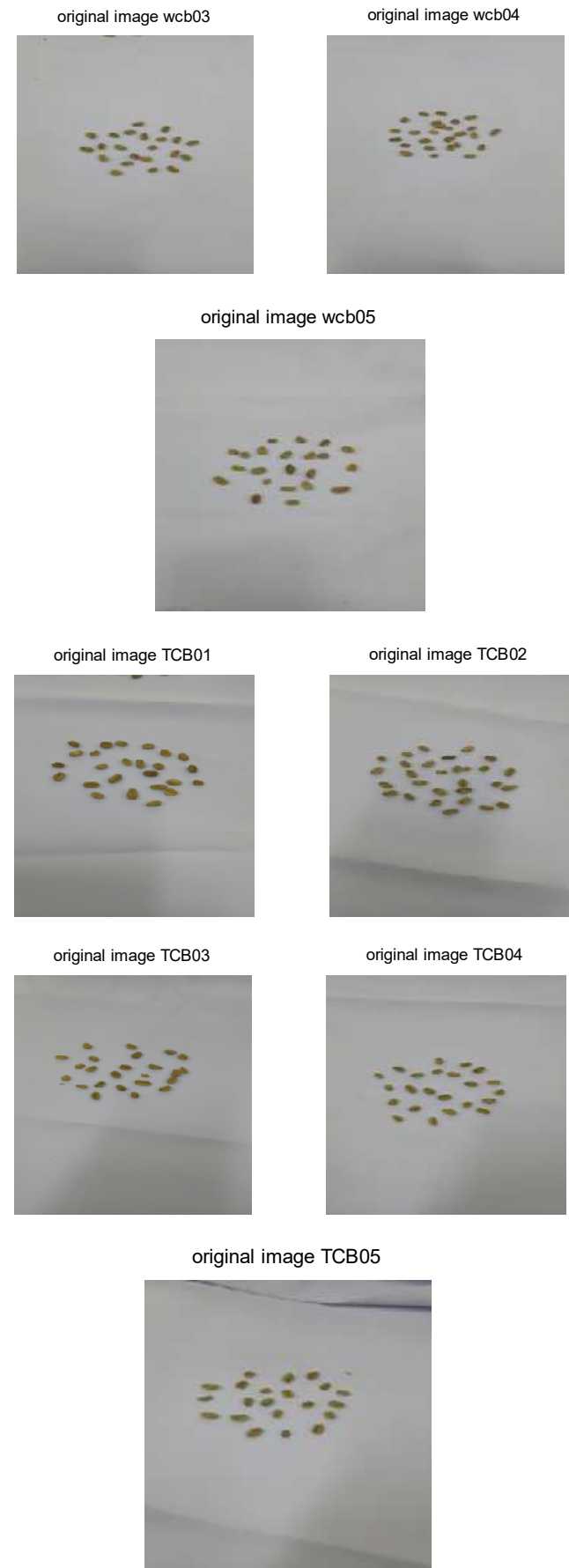


Figure 1. Original images of Wombera and Tongo coffee beans.

3. Methodology

3.1. Segmentation Result

Figure 2 and Figure 3 shows pre – processed images of the sample of coffee beans. The RGB images of sample coffee beans captured were converted into gray level images for feature extraction after the original images were enhanced by stretched contrast. Then the gray level image was changed to binary image by using automatic thresholding. The binary image indicated in the figure was obtained, but due to some defects on the surface of the coffee beans, there exists white color surface which resulted as hole. These holes were filled with their neighboring pixels, and then image of binary images with holes was filled as shown in the figure.

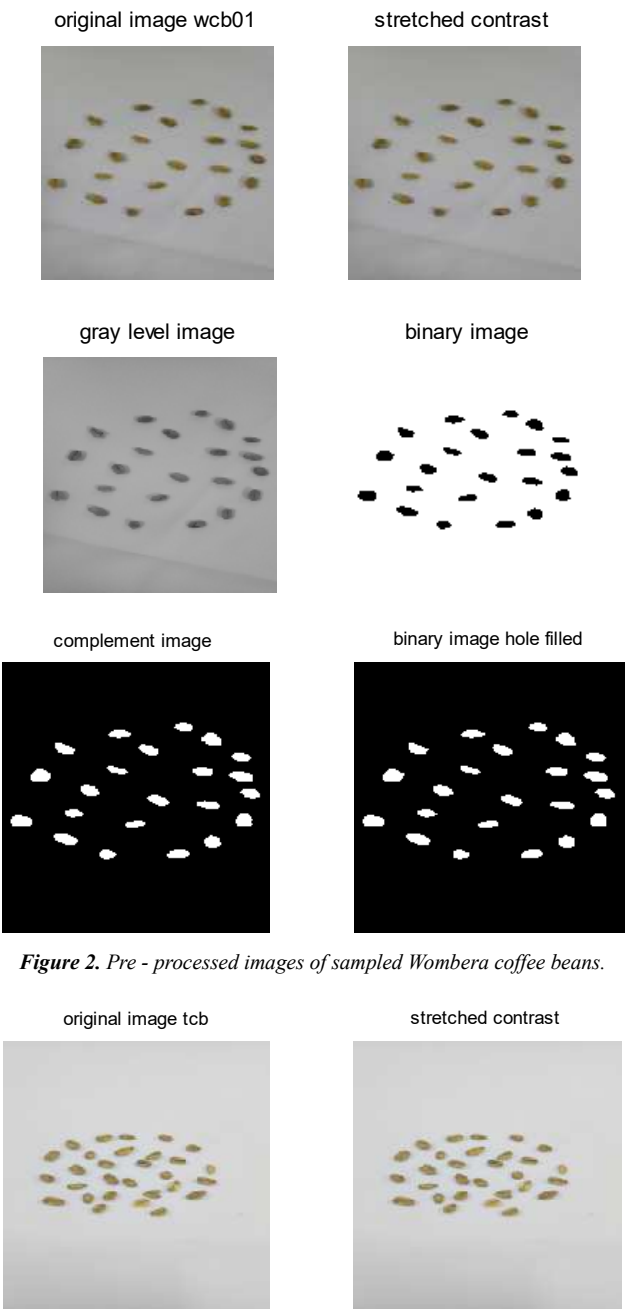


Figure 2. Pre - processed images of sampled Wombera coffee beans.

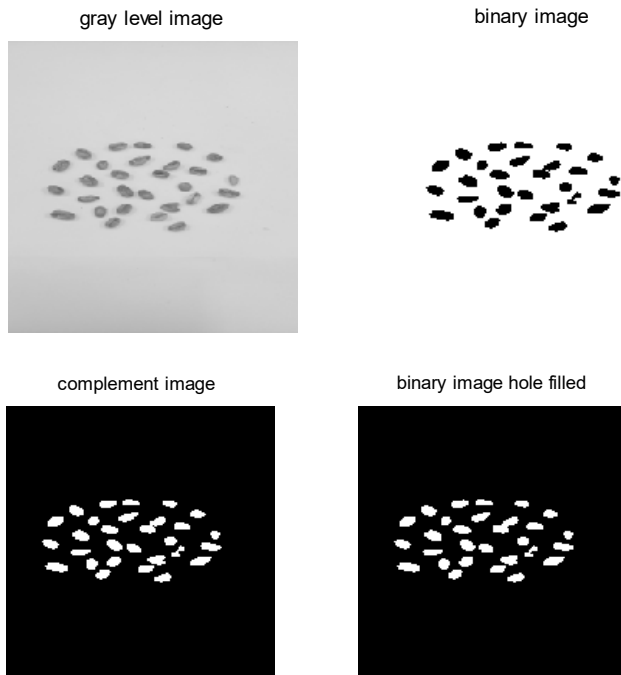


Figure 3. Pre - processed images of sampled Tongo coffee beans.

### 3.2. Result of Colour Features

Summarizations of the result of color feature are given in Table 1. In Table 1 Tongo coffee beans have the largest value of green and blue components compared to the coffee beans from other Benishangule coffee beans types, this gives Tongo coffee beans a bluish shade of green appearance.

Table 1. Sample colour features extracted from Benishangule coffee beans.

Feature Color	Benishangule coffee		
	Value	Wombera	Tongo
Mean	R	164.214	206.699
	G	164.121	206.132
	B	162.722	204.361
	H	0.191	0.206
Mean	S	0.0172	0.0211
	I	0.646	0.813
	H	0.0636	0.0793
	S	0.00411	0.00634
Variance	I	0.00125	0.00287
	H	0.958	0.967
	S	0.778	0.604
	I	0.459	0.471

### 3.3. Result of Texture Features

Texture features results are summarized in Table 2. Table 2 Wombera coffee beans has the largest value of energy and less value in entropy compared to the other Benishangul coffee beans type. This gives Wombera has the highest similarity and highest homogeneity compared to the other types of Benishangule coffee beans.

Table 2. Sample texture features extracted from Benishangule coffee beans.

Feature Texture	Benishangule coffee	
	Wombera	Tongo
Contrast	0.0384	0.0397
Correlation	0.9512	0.9541
Energy	0.4846	0.4378
Entropy	0.7579	0.8351
Homogeneity	0.9808	0.9806
Inverse	0.995129500	0.995124926

### 3.4. Result of Shape and Size Features

Shape and size features result are summarized in Table 3. In Table 3 Tongo were found to be the biggest in size where as Wombera were found to be the smallest in size and rounded in shape.

Table 3. Sample morphology features extracted from Beneshangul coffee beans.

Morphological Features	Benishangule coffee	
	Wombera	Tongo
Area	756.962000	802.468000
Perimeter	103.331400	106.105733
Maximum Diameter	37.653467	38.225533
Minimum Diameter	25.607000	26.766467
Equivalent Diameter	30.944933	31.891333
Surface Roundness	0.682400	0.703400

### 3.5. Classification Model Using ANN

#### 3.5.1. Classification Result of Color Features

For classification part of this work using color features, twelve features (Mean value of red, Mean value of green, Mean value of blue, Mean value of Hue, Mean value of saturation, Mean value of intensity, Range of hue, Range of saturation, Range of intensity, Variance of hue, Variance of saturation and Variance of intensity) are used as input and the target output is twelve, which are BGTCB1, BGTCB2, BGWCB1, BGWCB2 and BGWCB3.

To adjust classification set-up using color features twelve inputs, eighteen hidden layers and five output layers of neuron of particle recognition toolbox of MATLAB software of version R2013a and five outputs were used as described in Figure 4 below.

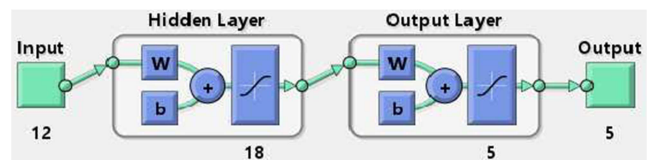


Figure 4. Network set-ups for classification using colour features.

The classification result based color features are shown in Table 4 of confusion matrix. The Table shows the confusion matrix that indicates the correct classification and misclassification of 100 sample images of the training 70% (70 images), validation 20% (20 images) and testing 10% (10 images) data.



**Table 4.** Confusion matrix result using colour features.

	BGTCB1	BGTCB2	BGWCB1	BGWCB2	BGWCB3	Total%
BGTCB1	40	5	0	1	0	87
BGTCB2	7	45	1	0	0	84
BGWCB1	0	0	48	0	0	100
BGWCB2	1	0	1	49	0	96.1
BGWCB3	2	0	0	0	50	96.2
Total%	80	90	96	98	100	92.8

As indicated in the Table the summary result of neural classifier using color feature alone showed that from the total of 100 images, 82 (82%) were correctly classified and 18 (18%) were incorrectly classified.

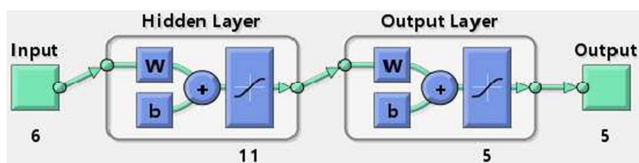
#### Analysis of the Result

As the result in neural network particle recognition toolbox (ANN) of overall confusion matrix of classification using color features indicate, that the classification accuracy of BGTCB1, BGTCB2, BGWCB1, BGWCB2 and BGWCB3 coffees were 80%, 90%, 96%, 98% and 100%, respectively. Only BGWCB3 was perfectly classified. BGTCB1 was more misclassified 7 image (7%) to BGTCB2, 1 image (1%) to BGWCB2 and 2 images (2%) to BGWCB3. BGTCB2 was more misclassified 5 images (5%) to BGTCB1. BGWCB1 was misclassified 1 image (1%) to BGTCB2 and 1 image (1%) to BGWCB2. BGWCB2 misclassified 1 image (1%) to BGTCB1.

From the result we could conclude that BGWCB3 has a unique color which makes it distinct from the others. There was also a strong color relationship between BGTCB1 and BGTCB2.

#### 3.5.2. Classification of Texture Feature

In this model, the six texture features (energy, contrast, entropy, homogeneity, correlation, inverse difference moment) of Beneshanguel coffee were used as input to the network. Hence, the neuron numbers of the input layer were six. The output neurons were five that correspond to the five predefined coffee growing regions considered in this study. The numbers of neurons in the hidden layers eleven as described in the Figure 5 below.



**Figure 5.** Network set-ups for classification using texture features.

The classification result based texture features are shown in Table 5 of confusion matrix. The Table shows the confusion matrix that indicates the correct classification and misclassification of 100 images of the training 70% (70 images), validation 25% (25 images) and testing 5% (5 images) data.

**Table 5.** Confusion matrix result using texture features.

	BGTCB1	BGTCB2	BGWCB1	BGWCB2	BGWCB3	Total%
BGTCB1	42	16	0	1	0	71.2
BGTCB2	2	34	0	0	0	94.4
BGWCB1	0	0	35	0	10	77.8
BGWCB2	6	0	2	35	0	81.4
BGWCB3	0	0	13	14	40	59.7
Total%	84	68	70	70	80	74.4

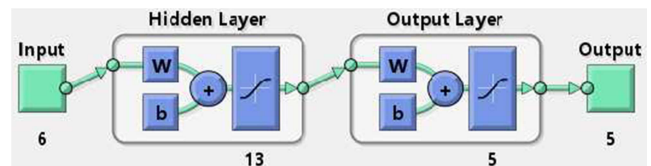
#### Analysis of the Result

The result of ANN classification using texture feature showed that the classification accuracy of BGTCB1, BGTCB2, BGWCB1, BGWCB2 and BGWCB3 were 84%, 68%, 70%, 70% and 80.0%, respectively BGTCB2, BGWCB1 and BGWCB2 were the more misclassified whereas BGTCB1 and BGWCB3 were more classified relative to the others. BGTCB1 misclassified 2 images (2%) to BGTCB2 and 6 images (6%) to BGWCB2. BGTCB2 misclassified 16 images (16%) to A. BGWCB1 misclassified 2 images (2%) to BGWCB2 and 13 images (13%) to BGWCB3. BGWCB2 misclassified 14 images (14%) to BGWCB3, 1 image (1%) to BGTCB1. BGWCB3 misclassified 10 images (10%) to BGWCB1.

From the result we could say that all Beneshanguel coffee samples share textural features, BGTCB1 with BGTCB2, BGWCB3 with BGWCB1 and BGWCB2, which may be attributed to the proximity of the region from which the coffee samples were drawn which possibly make them share certain genotype similarities. Resizing may be the cause because it may lose some basic information.

#### 3.5.3. Classification Result of Shape and Size Features

In this model, six shape and size features were selected namely, area, perimeter, maximum diameter, minimum diameter, equivalent diameter and Surface roundness. Hence, the neuron numbers of the input layer were six. The output neurons were five that correspond to the five predefined coffee growing regions considered in this study the number of neurons in the hidden layers were thirteen as shown in figure below.



**Figure 6.** Network set-ups for classification using shape and size features.

The classification result based shape and size features are shown in Table 6 of confusion matrix. The Table shows the confusion matrix that indicates the correct classification and misclassification of 100 images of the training 70% (70 images), validation 25% (25 images) and testing 5% (5 images) data.

**Table 6.** Confusion matrix result using shape and size features.

	BGTCB1	BGTCB2	BGWCB1	BGWCB2	BGWCB3	Total%
BGTCB1	0	3	0	0	50	94.3
BGTCB2	0	47	1	0	0	97.9
BGWCB1	2	0	40	0	0	95.2
BGWCB2	0	0	7	47	0	87
BGWCB3	48	0	2	3	0	90.6
Total%	96	94	80	94	100	92.8

As indicated in the Table the summary result of ANN classifier confusion matrix on the shape and size feature alone showed that from the total of 100 images, 82 images

(82%) were correctly classified and 18 (18%) were misclassified.

#### Analysis of the Result

The result of Artificial Neural Network (ANN) classification using shape and size features showed that the classification accuracy of BGTCB1, BGTCB2, BGWCB1, BGWCB2 and BGWCB3 coffee were 96%, 94%, 80%, 94% and 100%, respectively.

BGWCB3 was perfectly classified while BGTCB1, BGTCB2, BGWCB1 and BGWCB2 were misclassified. BGTCB2 misclassified 3 images (3%) to BGWCB3. BGWCB1 misclassified 7 images (7%) to BGWCB2, 1 image (1%) to BGTCB2 and 2 images (2%) to BGTCB1. BGWCB2 misclassified 3 images (3%) to BGTCB1. BGTCB1 misclassified 2 images (2%) to BGWCB1. From the result there is an existence of strong morphological relation between BGTCB1 & BGTCB2 and BGWCB2 & BGWCB3.

#### 3.5.4. Classification Model Result Based on All Features

In this model, twenty four features corresponding to six shape and size features, twelve color features and six textural features of Benshangul coffee were used as input to the neural network hence; there were twenty four neuron numbers for the input layer. The same to others, this experimentation has five output classes corresponding to the predefined coffee growing regions. The numbers of neurons in the hidden layers were also eighteen as shown in the Figure 7.

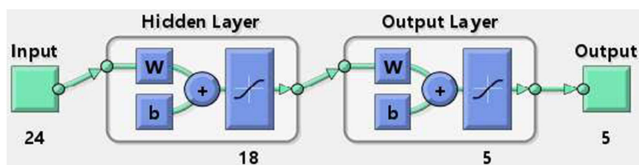


Figure 7. Network set-ups for classification using all features.

The classification result based on all features is shown in Table 7 of confusion matrix. The Table shows the confusion matrix that indicates the correct classification and misclassification of 100 images of the training 70% (70 images), validation 25% (25 images) and testing 5% (5 images) data.

Table 7. Confusion matrix results using all features.

	BGTCB1	BGTCB2	BGWCB1	BGWCB2	BGWCB3	Total%
BGTCB1	49	2	0	0	0	96.1
BGTCB2	1	48	0	0	0	98
BGWCB1	0	0	45	0	0	100
BGWCB2	0	0	1	50	0	98
BGWCB3	0	0	4	0	50	92.6
Total%	98	96	90	100	100	96.8

As indicated in the Table, the summary result of ANN classifier confusion matrix on all the features showed that from the total of 100 images, 92 images (92%) were correctly classified and 8 images (8%) were misclassified.

#### Analysis of the Result

The result of ANN classification using all features showed

that the classification accuracy of BGTCB1, BGTCB2, BGWCB1, BGWCB2 and BGWCB3 coffees were 98%, 90%, 96%, 100% and 100%, respectively. But BGTCB1, BGTCB2, and BGWCB1 were misclassified. BGTCB1 misclassified 1 image (1%) to BGTCB2. BGTCB2 misclassified 2 images (2%) to BGTCB1. BGWCB1 misclassified 1 image (1%) to BGWCB2 and 4 images (4%) to BGWCB3.

This shows that the classification result based on the combined features exhibited the highest accuracy and also there is strong relation between BGTCB1 and BGTCB1. There is a significant relation between BGWCB1 and BGWCB2. In other words, there is an ambiguity of classification between BGTCB1 & BGTCB2, and also between BGWCB1 & BGWCB3.

## 4. Conclusion

Coffee is a commercial commodity that plays a major role in earning foreign currency among export commodities of Ethiopia. The sub-sector is getting governmental and nongovernmental attention due its significance in commercial activities. The brand patent creation of each coffee variety based on region of origin was an issue in current period. Beneshangul coffee brands are not nationally (ECX) and internationally recognized and registered as property right to Ethiopia with their distinct character, flavor and taste. Beneshangul coffee beans which fulfill all the requirements especially Wombera coffee beans more power full than Wollega coffee beans in my researches and also traditionally very testy. The limitation of this area coffee beans are the researchers not attention to research this place so that the private and governmental organization with my study or additional research in other types of research methodology further research to recognize to get patent of this botanical coffee types.

The experimental results showed that color and morphology (shape and size) features have more accuracy to classify Beneshangul coffee based on growing regions than the other one feature (texture) in artificial neural network (ANN) classification. But the classification accuracy of coffee bean increases when all the features were used together, which shows the importance of incorporating all features in the classification Beneshangul coffees of different regions. Therefore, the coffee industry can use image processing as a quick way to identify the types of coffee beans.

## References

- [1] Andrea P., Svore, J. and Miertus, S. 1999, A biosensing method for detection of caffeine in coffee. *Journal of the science and Agriculture*, Vol. 7a, No. 12, p. 1136.
- [2] Endale Asfaw: *Physical Quality and Grading Systems of Ethiopian Coffee in Demand-Supply Chain*, Four Decades of Coffee Research and Development in Ethiopia 1967 -2007, 2007.

- [3] Ethiopia Coffee Quality Inspection & Auction Center: *Training Manual for Trainee Coffee Cuppers*, 2007. *manual, third edition*. 54-56 re de Montbrillant 1202 Geneva, Switzerland.
- [4] Faridah, G., Parikesit, F. and Ferdiansjah, 2011. Coffee bean grade determination based on image parameter. Department of Physics Engineering, Faculty of Engineering, Univertal Gadjah Mada Jalan Grafika 2 Yogyakarta, Indonesia. TELKOMNOKA, vol. 9, No. 3.
- [5] Furtado, J. J., cai, Z. and Liu, X. 2010. Digital image processing: supervised classification using genetic algorithm in matlab toolbox. China university of geosciences, 388 LuMo road, Wuhan, Hubei, P. R. china Zip code 430074.
- [6] Gonzalez, R. C. and Woods, R. E., 2002. *Digital image processing second edition*. Med Data Interactive. University of Tennessee.
- [7] Habtamu Minassie, 2008. Image Analysis for Ethiopia Coffee Classification. A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment for the Degree of Master of Science in Computer Science Addis Ababa University School of Graduate studies, Addis Ababa, Ethiopia.
- [8] Hammouda, I. and Rudzki, J. 2004. Pattern classification. Tempere University of Technology, institution software systems, 8109103-ohjelmistotuotannon Teoria.
- [9] Huiyu Z., Wu, J. &Zhang, J. 2011. Digital image processing part I. Ventus publishing aps, ISBN 978-87-7681-541-4.
- [10] *International coffee organization*, 2003. Coffee and health, new research finding proceeding of the international seminar on coffee and health 40<sup>th</sup> Anniversary meeting of the ICO Cartagena, Colombia.
- [11] International trade centre, 2011. *The coffee exporter's guide*
- [12] International Trade Center UNCTAD/WTO: *Coffee An export Guide*, Geneva, 2002.
- [13] Maheshwari C. V., Jain, K. R. and Modi, C., 2012. Non-destructive quality analysis of Indian Gujarat-17 Oryza sativa SSP Indian (Rice) using image processing. *International Journal of computer engineering science (IJCES)*, volume 2 Issue 3.
- [14] Nicolescu, C. and Jonker, P. 2002. A Data and Task Parallel Image Processing environment. Elsevier Science B. V., Parallel Computing 28945-965.
- [15] Pathon M., Jussoff, K. and Barkatullah, Q., 2011. Implication of image processing algorithm in remote sensing and GIS application. *Journal of theoretical and applied information technology* volume 34 No-1.
- [16] Seeman, T., 2002. Digital Image Processing using Local Segmentation. Submission for the degree of Doctor of Philosophy, School of computer science and software Engineering. Monash university, Australia.
- [17] Surendra Kotecha and Ann Gray: *ICO/CFC Study of Marketing and Trading Polices and Systems in Selected Coffee producing countries: Ethiopia Country Profile*, 2000.
- [18] Unay, D. and Bernard, G., 2005. Artificial Neural Network-baTCTS Labs., Faculty Polytechnique de Mons Multitel Building, Avenue Copernic 1, Parc Initialis, B-7000, Mons, Belgium.
- [19] Waller, J. M., M. Bigger and R. J. Hillocks: *Coffee Pests, Diseases and their Management*, Column designs Ltd, UK, 2007.